University of Canterbury

Human Interface Technology Laboratory
New Zealand

Ego- and Exocentric interaction methods for mobile AR conferencing

Author:
Timo Bleeker

Supervisor:
Prof. Mark Billinghurst
Co-Supervisor:
Dr. Gun Lee

Dissertation submitted in partial fulfilment for the degree of
Master of Human Interface Technology at the
Human Interface Technology Laboratory New Zealand,
College of Engineering,
University of Canterbury
July 7, 2013

Acknowledgments

The author wishes to express sincere appreciation to project supervisor Professor Billinghurst for his fantastic guidance and supervision, and for being a source of inspiration. The author also would like to thank secondary supervisor Dr. Lee for all his help and valuable support with analysing the study data. Many thanks go out to Samuel Williams for the enlightening guidance he has provided for programming the application. In addition, the author would also like to thank all students and staff of the Human Interface Technology Laboratory New Zealand for all support and inspiration, as well as creating a comfortable and enjoyable research environment.

Abstract

Augmented Reality is technology that superimposes virtual content on the real world, typically shown through a see-through head mounted display (HMD) or handheld device. AR has successfully been used for many applications and provides new opportunities for remote collaboration and communication. With the growing availability of commercial HMDs such as Google Glass and the Oculus Rift, more possibilities in the field of AR have opened up. However, interaction with AR content shown on HMDs is still not very well explored. This master's thesis investigates the possibilities of a combined use of head mounted and hand held displays (HHD) for interaction in AR conferencing experiences. Prior research in communication, AR collaboration and HMD-HHD interaction is reviewed before presenting new interaction methods. Two different HHD interfaces and cuing methods were created to support file sharing in an AR conferencing application. A formal evaluation compared four different combinations of the interfaces and cuing methods. The results showed a significant difference between the different conditions where in particular one condition performed better than the others. The results were used to create a set of basic design guidelines for future research and application development.

Contents

List of Figures

1 Introduction

This thesis explores how a hand held display can be used for input in an Augmented Reality conferencing application shown on a head mounted display. Augmented Reality (AR) is technology that allows virtual imagery to be overlaid on the real world. It has been used in many different types of applications, such as education, engineering and entertainment. Many of these are single user experiences, but there has also been research conducted on how AR can be used to provide new types of collaborative experiences.

Over the last decade smart phones and other hand held devices (HHD) have become very popular, and now provide a powerful platform for mobile computing and communication. Similarly, consumer HMDs that will provide Augmented Reality and Virtual Reality experiences, such as Google Glass [16] (see Figure 1), and the Oculus Rift [32], are entering the market. However, there is a need for new ways of interacting with these devices. For example, although people can interact with Google Glass using speech and touch pad input, more precise input might be required for more complex applications. Since Google Glass can already be connected to a mobile phone, this phone could also be used for HHD input. Hybrid interactions that span multiple devices may be very well suited for HMDs. However only a relatively small amount of research has been done on the combined use of head-mounted and hand-held displays, especially for AR applications.

Figure 1: Google Glass

In this research we are particularly interested in how hybrid HMD and hand-held devices can be used to support remote collaboration. Previous research has shown that head mounted displays can be used to create mobile AR collaboration spaces that use spatial cues for enhanced communication [14] [26]. However these systems just used simple input techniques previously available, such as a mouse and keyboard. Combining a hand-held display with a head mounted display could lead to more intuitive input methods for mobile AR conferencing applications.

This thesis will explore several interaction methods in an AR conferencing setting, with the intention of determining fundamental design guidelines for interaction between a hand-held and head-mounted display. While there has been previous work on the combined use of hand-held and head mounted displays, there has been little work focusing on remote collaboration. This work tries to close this research gap and will provide a basis for future work on similar applications.

Inspiration for this project came from some of the author's own experiences with remote collaboration. Several projects have required the use of remote conferencing software to collaborate with team members. For example, using instant messaging or video calling software. However, while both have their strong points, neither of them seems as effective as being there in person. Often the author was talking to a group of people who were physically together, while he was the remote team member. Even though communication often went better and more streamlined than expected, the author often felt disconnected from the conversation. As a remote user, it was difficult to see the body language of all the participants, nor could the remote participants see all of the author's actions. This made it very difficult to participate in the discussion. This personal experience helped the author to develop the interaction concepts for remote collaboration applications described in this thesis.

In order to explore how a HMD and HHD can be used together in AR conferencing, an iterative design process was applied (See Figure 2). This starts with investigating and identifying the problem, and conducting a user needs analysis to determine key user requirements. Next, possible solutions or products are designed and planned. After the planning phase the solutions can be created in either low or high fidelity forms to finally evaluate them in the last phase of the Design Cycle. The data gathered in an evaluation is then used to repeat the cycle over again with a better knowledge of the problem area. Generally, a first iteration of the Design Cycle can be done in hours. With every iteration the

time taken for each phase grows, the problem is understood better and solutions are created with higher and higher fidelity.

Figure 2: A representation of the Design Cycle

The thesis is broken into several chapters following the basic structure of the Design Cycle: investigation, planning, creation and evaluation. First, the chapter *Related Work* summarizes and reviews previous related work in the areas of communication, virtual conferencing, spatial audio, collaborative AR and examples of the combined use of a head-mounted and hand-held display. There will also be a short review on relevant evaluation methods, and a summary of research opportunities. This chapter mainly covers the Investigation and touches the Planning phase with the research opportunities section.

The next chapter *Design Process* will describe the design process that resulted in the final concept and user experience design. This chapter will describe the user centred design process followed, the results of initial brainstorming, a user and use case, and what external material functioned as inspiration. The chapter will close with a full description of the concept and interaction methods. The Planning phase of the Design Cycle is represented by this chapter, with some elements of the Creation and Evaluation phases.

Following the design process, the chapter *Implementation and Prototype* will focus on the hardware and software development of the application. It will review the available technology to support the final hardware decisions. and show the development process of the software, both on Windows and Android platforms. This chapter represents the Creation phase of the process with some references to the Evaluation and Planning phases.

The chapter *User Evaluation* will explain a user study conducted to evaluate the prototype design. It provides an overview of the study design and procedures, as well as describing the participants demographics. The *results* section describes the statistical tests and results that have been discovered in the study. The end of this chapter will interpret the results and draw conclusions. The last element in the Design Cycle, Evaluation, is covered by this chapter.

The *Discussion and Design Guidelines* chapter provides an in-depth discussion of the results found, and from this describe a set of design guidelines that could be used by others to build similar interfaces. This chapter lays the foundation for a following iteration of the Design Cycle by touching the Evaluation, Investigation and Planning phase.

The final chapter *Conclusions and Future Work* will conclude the thesis with a summary of the whole project and all iterations, suggest future work and describes the lessons learnt during the project.

The main contributions from the thesis are as follows:

- An exploration of Exocentric and Egocentric interfaces for collaborative HMG-HHD interaction.
- An evaluation of the effectiveness of Spatial Sound cues versus Visual cues for data sharing in collaborative interfaces.
- Provision of a set of design guidelines for future HMD-HHD AR conferencing applications.

2 Related Work

As discussed in the Introduction, this thesis explores how a hand-held display can be used in combination with a head-mounted display to provide innovative and interesting interaction methods for use in an Augmented Reality conferencing space. This chapter of the thesis will review relevant previous research work, divided into several sections. The first few sections on *Communication*, *Virtual Conferencing* and *Spatial Audio* review work mainly focused around the communication part of (remote) collaboration. Next, the sections on *Collaboration with AR* and *Head Mounted and Hand Held Displays* review work about interaction and interfaces for AR. The final section, *User Evaluation*, will summarize relevant research on evaluation methods. The chapter closes with a summary describing the research opportunities and gaps, and how our work addresses those gaps.

2.1 Communication

Communication is an integral part of collaboration, whether it is local face-to-face communication or communication via a remote connection. However, remote collaboration is often hindered by the lack of conversational grounding and situation awareness. Conversational grounding refers to the mutual understanding of subject in a conversation and grounding techniques change with the communication medium used. Each form of medium might impose different constraints [4]. For example face-to-face communication is instantaneous while a letter will take some time to arrive to the addressed person. However, reviewing face-to-face conversation might be nearly impossible while a letter can be read over and over again before a reply is send.

Situation awareness is a term used for people's mental models of complex, dynamic environments [7]. This means that people are aware of the situation around them, understand it and are able to intervene to work towards a desirable future outcome. For instance, in a collaboration between a student and mentor, situation awareness can refer to the awareness the mentor has of the required amount of help at a certain time. Collaboration is generally more efficient and simplified when both conversational grounding and situation awareness have been established. One way this could be achieved is by streaming video (Figure 3) of a worker's task space to a remote expert, who is then able to better instruct the worker on how to complete the task [18].

Previous research has shown that some media does not have any effect on task completion time in a remote collaboration task, but they do change the

Figure 3: Remote Collaboration with video streaming

way people communicate. For example, a video connection between two remote people allows them to communicate with fewer descriptive words and more with gestures [19]. This is because fewer words are needed to establish grounding and create good shared situational awareness. Video communication can be further enhanced by allowing people to draw gestures on the shared video feed (Figure 4), and so use an additional communication cue for grounding. Users also tend to use hand drawings to illustrate movements and angles of insertion [9].

Figure 4: Drawings on a remote user's view

2.2 Virtual Conferencing

Talking on the phone is different from a face-to-face conversation. A lack of body language and other visual cues removes the ability for people to gain common conversational grounding. However several studies have shown that remote

conferencing can be improved with Augmented Reality and Virtual Reality techniques.

In Shared Virtual Environments remote users come together in a virtual space where they are represented by virtual characters or avatars. For example, each user could be represented in the virtual environment by a small window containing a live video feed of their faces. This window also functions as the user's viewport into the virtual environment (Figure 5). As a result, users become aware of each other's gaze and location in the virtual environment and have an increased sense of Social Presence and spatial presence when collaborating with each other [14].

Figure 5: User's views from the virtual room view

In a comparative study where a virtual conferencing system was compared to more traditional conferencing methods, it was found that the virtual conferencing room had a slower task completion time. However, it was also found that the 3D virtual conferencing system supported more spatial cues such as gaze awareness than the tested 2D interface, producing a higher social presence and co-presence [12]. This extended an earlier study comparing 2D and 3D collaboration spaces that also found that a 3D collaborative interface invokes a stronger feeling of Social Presence than a 2D interface. However both 2D and 3D interfaces score lower than face to face collaboration [11]. These results show that 3D virtual conferencing systems can introduce spatial cues that create a stronger feeling of Presence.

2.3 Spatial Audio

Another important factor of virtual conferencing is the ability to discriminate between several simultaneous speakers. Spatial audio can be used to identify and navigating several different audio streams and differentiate between speakers [2]. The use of spatial audio also supports background awareness and focus direction [5], which makes it a viable implementation in virtual conferencing.

This assumption is further strengthened by a study in which the effects of spatial audio on memory and comprehension in desktop remote conferencing was tested [1]. Users were significantly better at remembering and identifying speakers when spatial audio was used, compared to a condition without spatial audio.

2.4 Collaboration with AR

When a task at hand is too complex for one worker to execute, assistance by a remote expert can significantly decrease task completion time and increase the quality of performance [19].

Several systems have explored the use of Augmented Reality for enhancing remote collaboration. For example, in the C-Slate system [17] two users were able to work on a shared space on individual tablets. The user's hands are tracked and shown as an overlay on the remote display allowing for precise collaboration or teaching (see Figure 6). Although no formal user study has been performed yet, people have expressed positive feelings about how C-Slate offers a natural and expressive way of remote collaboration.

Figure 6: Remote user's hands shown on local user's display

AR annotation by a remote expert using projectors is another way of supporting collaboration. Annotations can be projected directly onto the object of interest giving the local worker a display-free experience and providing a very direct method of supporting the user [23]. Unfortunately, projectors have several disadvantages when used for annotation projection. It needs to be very bright, such as with a laser projector, in order to be used in typical work environments and although the projector doesn't require to be operated by the local worked on the scene, it has to be carried there and placed and powered on location.

An alternative to annotation projection is the use of Head Mounted Displays (HMD). An HMD is personal, portable and also leaves the worker's hand free for any other tasks. A combination of both a projector and see-through HMD is used in the Studierstube system [27]. This system was tested with a story-boarding application, where two users could remotely collaborate on the same project. The system used three display methods of which two were Augmented Reality displays and one was projected on the wall. The AR views were projected on the Personal Interaction Panel (PIP) (see Figure 7) and appeared as a floating box just in front of the user. The PIP allowed users to interact with the scenes inside the floating boxes. The wall projection was a display shared between the users and showed details of a particular scene. With this system users were able to work on separate tasks using the HMD while working together on the projected surface.

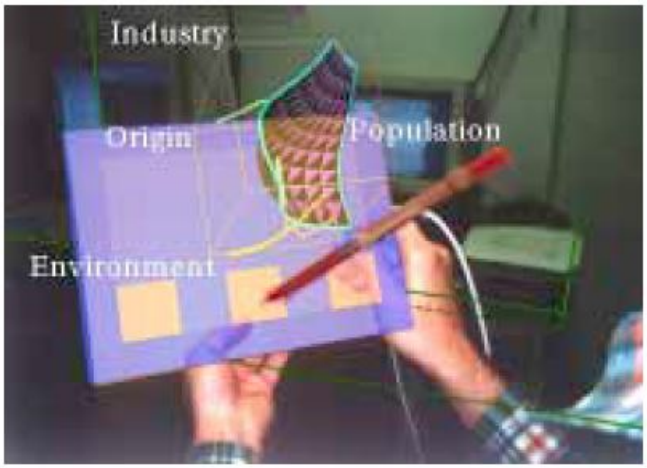


Figure 7: A Personal Interaction Panel (PIP)

Another example of using an HMD for remote collaboration is the AR system for use in a crime scene investigation described in [24]. In this case, the local user wears a head mounted display and stereo head mounted cameras. A remote expert can place virtual annotations at points on the image the local user sees. A head worn stereo camera created dense point-clouds used to make a 3D model of the environment. These cameras were also used to tracked the user's hand gestures for input. The remote expert also used these cameras to see the scene just as the worker was seeing it (see Figure 8). In this way users did not need a lot of communication to acquire a mutual understanding.

The main problem encountered in this system however, was the lack of vir-

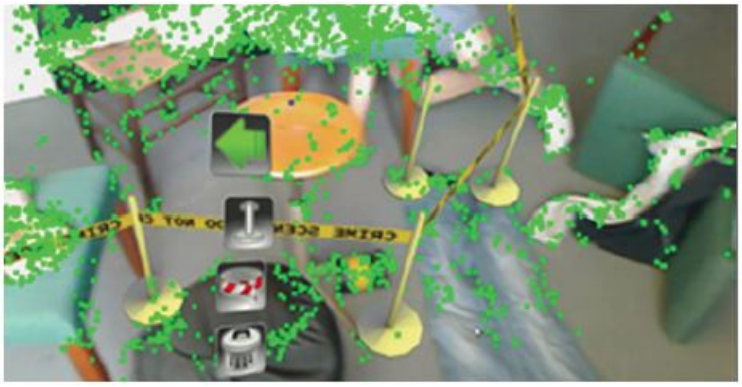


Figure 8: 3D feature map and AR objects overlaid on the camera view

tual co-presence. This made it difficult for users to effectively communicate, with both users often talking at the same time or interrupting each other. There were also other issues, such as slow response time of the pose tracker, which restricted the local user to slow movements. Conversational disconnection also occurred between participants when the remote expert placed or moved objects in the scene without verbal indications.

As these systems show, AR provides the ability to create virtual annotations over a remote user's view and also share gestural cues that can significant help the remote collaboration. In the next section we describe in more detail how HMD and handheld displays are used in AR systems.

2.5 Head Mounted and Hand-held Displays

Head mounted displays have been used since the very first AR systems in the 1960's to allow virtual graphics to be superimposed over the real world [28]. However, while a HMD is portable and hands-free, it often lacks display brightness and resolution. There is also a need to support different input methods for the AR content seen in the HMD. One way of doing this is through adding a hand-held display that can act as a secondary high-resolution screen that also enables input through touch or using several embedded sensors.

One of the first systems in which a HMD and HHD were combined were the Touring Machine [8] and the MARS system [15] developed at the University of Columbia. In these systems, multiple display and interaction technologies and their complementary capabilities were tested. Users of the Touring machine wore a HMD and carried a HHD (see Figure 9). The Touring Machine was used to show information about buildings around the user. While looking through

the HMD, significant buildings were labeled with virtual information. Using the orientation tracker of the HMD, users could look around. When a label would come closer to the centre of the screen, it would increase in brightness. As soon as it reached the centre of the screen, it would turn yellow and green, indicating selection. The HHD showed additional information about the buildings and the information was navigated through using a stylus on the touch screen. When a label is selected from the HHD, it is automatically selected on the HMD.

Figure 9: The MARS system used for the Touring Machine

The MARS system is very similar to the Touring Machine, but additionally offers the ability of communication between outdoor and indoor users. The indoor user can get an overview of the outdoor scene and communicate with outdoor users through an immersive AR user interface (see Figure 10).

Figure 10: AR interface for the indoor user of the MARS system

Another project that combined multiple display technologies was EMMIE [3]. In this system items were presented either on a HMD, a tablet display, a laptop or a projected surface (see Figure 11). This project used hybrid interaction, to manipulate data in the system and so users could drag virtual objects from one display and drop them onto another display. The system also used a tracked display to view otherwise invisible information.

Figure 11: A meeting situation using EMMIE.

An interesting use of an HHD is shown in ARWand [10], where the internal orientation sensor of a smartphone was used to manipulate 3D objects in Augmented Reality. The user could point at AR objects with the phone and then manipulate them by tilting the phone (see Figure 12). However, no usability study was conducted to evaluate the system.

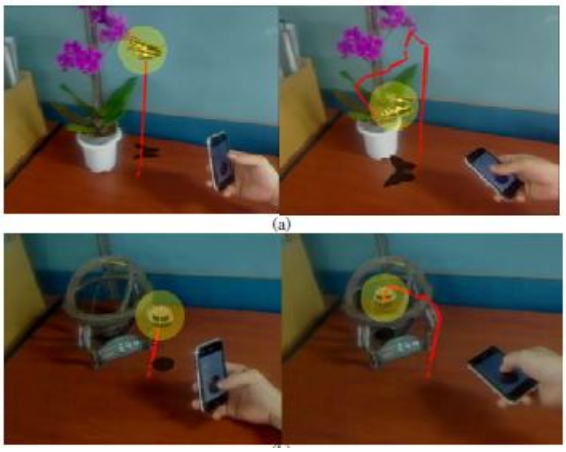


Figure 12: The ARWand used to control and AR scene

A mobile display is also used to control items on an HMD in [13]. The project evaluated Augmented Reality for routine maintenance tasks inside armoured vehicles. The head-mounted display showed the user a 3D animation of a certain task. A wrist mounted touch-enabled phone was used to control the animations. The phone had a simple user interface with a slider bar to control the speed of the animation, forward and back buttons to navigate between different tasks and a stop and start button to pause and resume the animations respectively (Figure 13).

Figure 13: Wrist mounted display used as controller

While not actually using a hand held display, the work of Szlavári and Gervautz [29] does explore interaction methods for the combined use of a handheld input device and HMD. Instead of a real display, the system uses a 'dumb' panel and pen with a see-through head mounted display (Figure 14). One of the reasons for using a dumb panel over a HHD was the physical limitations of existing HHD devices such as size and weight. The work explores interesting interaction methods for object manipulation, navigation, system control and sharing. However, the dumb panel also has its limitations. Interaction with the AR scene has to be done through the use of a tracked pen, which limits the user's ability to use their hands to manipulate real world objects. The panel also cannot provide a high-resolution secondary display.

In summary, it can be seen that there have been a number of research efforts exploring how hand held displays and input devices can be used with HMD

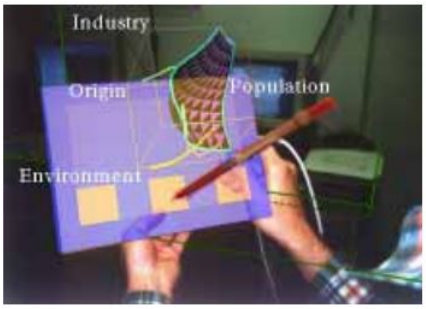


Figure 14: AR images projected on the tracked 'dumb' panel seen through an HMD

for AR systems. However these systems have mostly been focusing on how to support interaction with virtual content in the AR scene, and not how to support collaboration. In our research we are interested in exploring how HHD and HMD can be used together to enhance remote AR conferencing and collaboration.

2.6 User Evaluation

In addition to developing a prototype HHD-HMD system for AR conferencing we also want to conduct a user evaluation of the system. This section summarizes some related work done on evaluation methods that could be used to evaluate the final prototype.

The Post-Study System Usability Questionnaire created by IBM [20] records participant answers in three different factors; System Usefulness, Information Quality, and Interface Quality. Although this questionnaire has limited generalizability due to a relatively small sample size, it provides a good base to built a questionnaire on for this research.

One important factor of remote collaboration and communication is the feeling of Presence which is connected to the user's feelings of involvement and immersion. The questionnaires described in [33] test for these feelings. They have been evaluated in a real and virtual situation [31] and while the expected significant difference was not found, the questionnaire can still be used to compare different interaction methods within virtual environments.

Billinghurst, Dünser and Grasset surveyed evaluation techniques used in AR studies and found five types of evaluations used: Objective measurements, Subjective measurements, Qualitative analysis, Usability evaluation techniques and Informal evaluations [6]. They found that most AR studies use Objective

measurements such as task completion times and accuracy (75 publications) but only 7 publications employed Usability evaluation techniques. This suggests that most AR studies focus on a very technical aspect of AR applications and do not directly study usability and user experience.

2.7 Research Opportunities

In summary, in this chapter we have shown that interesting research has been conducted in the areas of interaction design for HMD and HHD usage, as well as for remote collaboration. However little research has been done on the use of a HHD as an extended interaction platform for an HMD. With head mounted displays such as Google Glass soon becoming commercial products, more exploration needs to be done in interface and interaction design for these devices. One important research gap here is the lack of comparative studies between different interfaces for the combined use of an HMD and HHD.

Quite a lot of research has been done on remote collaboration applications with and without using AR. However, there is still a research gap in the interaction design for remote collaboration using AR. There are several systems that explore interaction methods for (remote) collaboration with AR, but there are no clear guidelines in place for other research or commercial applications to build on. There are still many opportunities to design and test systems for real world situations.

Our research is the first to explore the use of head mounted and hand held displays for AR conferencing and aims to fill above mentioned research gaps. In this way we hope to create a foundation for other research in this area.

3 Design Process

This chapter describes the Planning phase of the Design Cycle, in which possible solutions are designed. It also shows how initial ideas were created and evaluated in short iterations until the final concept was created.

3.1 Inspiration and Idea Generation

This section deals with the iterative design process that preceded the development of the final prototype. It starts by describing the brainstorm session that was held to generate initial ideas. The next section deals with the user and use case, which were used as guidelines to further develop the ideas generated in the brainstorm session. Finally other inspiration and influences are also discussed.

3.1.1 Brainstorm

In preparation for the brainstorm session, a discussion was held to find a common understanding of the qualities and limitations of head mounted and hand held displays. The assumptions agreed upon are described below. These are assumptions made to find a common ground and simplify the brainstorming process.

Assumption 1: A head mounted display is personal and intimate. Since it is worn on the head it shows information only to the wearer. This suggests a use for private and personal information shown only to the wearer. However, in case of a see-through display, the information shown on the HMD is superimposed on top of the public world. Items and information can appear on the display as Augmented Reality. This also allows for a common augmented world that can be seen and interacted with by more than one person at the same time for collaboration. While the physical display on an HMD is relatively small the effective display can be all around the user when the head orientation is tracked. An HMD can be always-on and show notifications when needed. However when an HMD display is cluttered it can obstruct the user's vision [34].

Assumption 2: A hand held display is less intimate than a head mounted display. It is portable and may have a large high-resolution screen that is touch enabled. It may also have several sensors for orientation, a camera and wireless connectivity. The physical display may be larger than the physical HMD display as, the effective display ends at the boarders of the screen.

Together with other researchers at the HITLabNZ and with the above as-

sumptions in mind, a large number of ideas were generated during a brainstorm session at the beginning of the project. The ideas were all focused around remote collaboration with AR and the use of a head mounted display with a hand held display. In five minutes as many sketches and ideas as possible were created. When the five minutes were over, each person was asked to pick two of their favourite ideas and explain them to the group. While explaining the ideas others could also give feedback and elaborate on the idea. Afterwards the ideas were categorized into two categories: communication and interaction. All ideas are shown below. Ideas that helped create the final concept are elaborated on.

Communication

The following ideas were categorized under Communication. Some ideas that were relevant to the final concept are further explained below.

1. Use HHD as second camera - show view of person or tight space.
2. Use HHD to capture scenes and objects to create 3D models.
3. Peer to peer remote collaboration - can share video from other persons HMD on HHD and annotate on the view and share back, so person can see annotation of their own view in their own HMD.
4. God like navigation method - outside user sees virtual hand in real world for navigation.
5. Any communication can be recorded when virtual object is selected for training or reviewing purposes.
6. Use HHD for private space and HMD is used for showing public information.
7. Use HHD for asynchronous communication - monitoring remote people - flick people into HMD view to start synchronous conference.
8. Remote AR people for conferencing - can share documents using HHD and gestures.

One of the ideas that inspired the research further was about the possibility of dedicating a virtual space to each of the display types. In a virtual environment, a hand held tablet could function as a personal space where documents are kept only for the local user. The HMD could show other (remote) participants and a virtual table or designated area for public files and interactions (see left part of Figure 15).

An HMD could be used to show a virtual conferencing room, while a hand held display could be used to interact with the participants in the conference. For instance, a tablet could be used to make a selection of people and place them into the virtual conferencing space shown on the HMD. Participants could also be relocated or removed from the HMD by using the tablet (see right part of Figure 15).

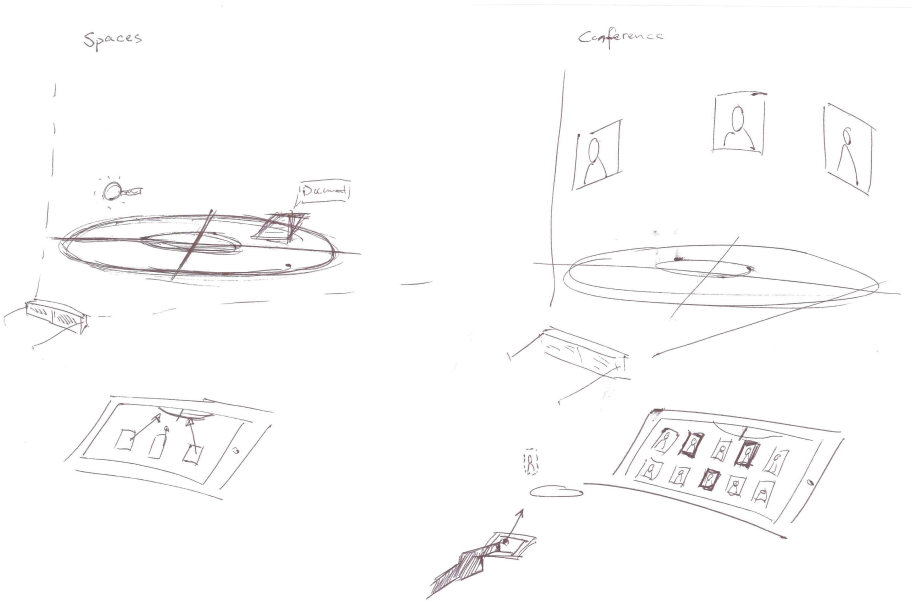


Figure 15: Left: virtual spaces. Right: virtual conference.

Interaction

The following ideas were categorized under Interaction. Some ideas that were relevant to the final concept are further explained below.

1. Use HHD to manipulate virtual objects.
2. Use HMD as secondary display for HHD - pin information from HHD to HMD.
3. Use HHD to capture scenes and create 3D objects.
4. Use HHD as lens for HMD to provide higher resolution part of the image/real world.
5. Minimal attention display - just show alerts - alert with HHD for people to put on the glasses.

6. Track HHD and use this for interaction; create virtual conferencing space by using tablet to fix people in space.Can use tablet for writing and adding annotations.

7. Use HMD + HMC to capture user actions then tablet will capture user actions and make dynamic manual.

8. Camera on HMD can zoom depending on position of tracked camera.

9. Use HHD for manual - flick out portions of manual and see AR highlights appearing on HMD.

10. Use HHD and HMD sensors for tracking and HHD for annotating real world and annotation can be viewed in HMD.

There are not many ways of directly interacting with data on a HMD. A hand held display can help with interaction and can be used to manipulate the data shown on the HMD. This could mean data creation, positioning and deletion etc. For example, there are several ways a virtual conference call could be initiated. A user could simply tap the people required for the conference, or drag them to a certain area on the HHD. Other, more novel interactions have also been suggested. A user could for instance position virtual participants in a AR conferencing space by holding up the tablet in front of them to place a virtual participant at that location, resembling the action of attaching photo frames to the wall (see Figure 16). Or a user could swipe from the tablet in a certain direction to throw the participant there, similar to dealing a deck of cards.

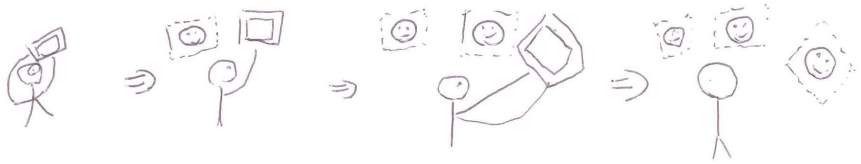


Figure 16: concept of interaction for initiating a conference call

A hand held display can show a lot of information at the same time because of its large high-resolution screen. However, having the same amount of information on a HMD is inconvenient and may cause confusion or clutter. By selecting certain elements on the HHD and pinning them to a HMD, a user could decide what information is useful at that moment. For instance an image or piece of text could be virtually pinned to a real-life object. Interaction could be enhanced by exploring different methods of pinning information from the

HHD to the HMD. One suggested idea was to flick the information objects from the HHD to the HMD.

3.1.2 User and use case

Initially, the project was focussed on remote collaboration using AR and a HMD in combination with a HHD. The target user in this scenario is a worker who is required to work in a remote location and maintain and repair complex systems of which the workers knowledge might be limited (see Figure 17). In this case, workers often request support from remote experts to evaluate the problem and decide what actions to take next. Currently, support is given either just by phone, with occasional pictures being send back and forth, or by having an expert come to location to take a closer look at the problem, which can be very time and resource consuming. In contrast, Augmented Reality could be used to enable a remote expert to annotate the workers field of view with virtual cues that could help them perform the tasks. This may improve the collaboration and reduce the need for the remote expert to travel to the worker's location.

Figure 17: An example of a remote worker and work environment (Credit: Siemens press picture)

Tasks that users of such a system might want to accomplish are

- Exchange photo's, manuals or other documents
- Share live video stream of the work environment
- Consult with multiple people
- Receive live AR annotations
- Look up (online) resources

A system for this environment also has some physical requirements:

- Lightweight and portable
- Not obstructing the user's natural movement and vision
- Easy to learn and use
- Useful

There is a relatively large field of research dedicated to virtual conferencing, and many aspects have been studied before. Therefore the thesis research was focused on a small part of this greater research area. In particular we were interested in the problem of how to share content between participants in an AR conferencing application. One of the goals was to find a way to make the sharing of files between conference participants as fast and effortless as possible.

We imagine an AR conferencing application where a typical user would be somebody in a remote location, who is in need of assistance from one or more remote experts or advisers. In a scenario like this, the user would initiate a conference call with the required people. A virtual space would be shown in the HMD, and the contacted people appear as virtual images in the space. The user could position the people as desired, so that their virtual images appear spread around the user in the real world.

With the help of spatial audio, it would be easy for the user to locate the person that is currently talking. During the conference call a participant might require the user to send a file, or vice versa. Spatial audio could be used to make this request clear. For example, in case of remote collaboration somebody might require the user to send a photo of a broken piece of equipment to identify it. The same person could also share a 3D model of the specific part with all participants and continue the conversation from there. Figure 18 shows a sketch of the concept.

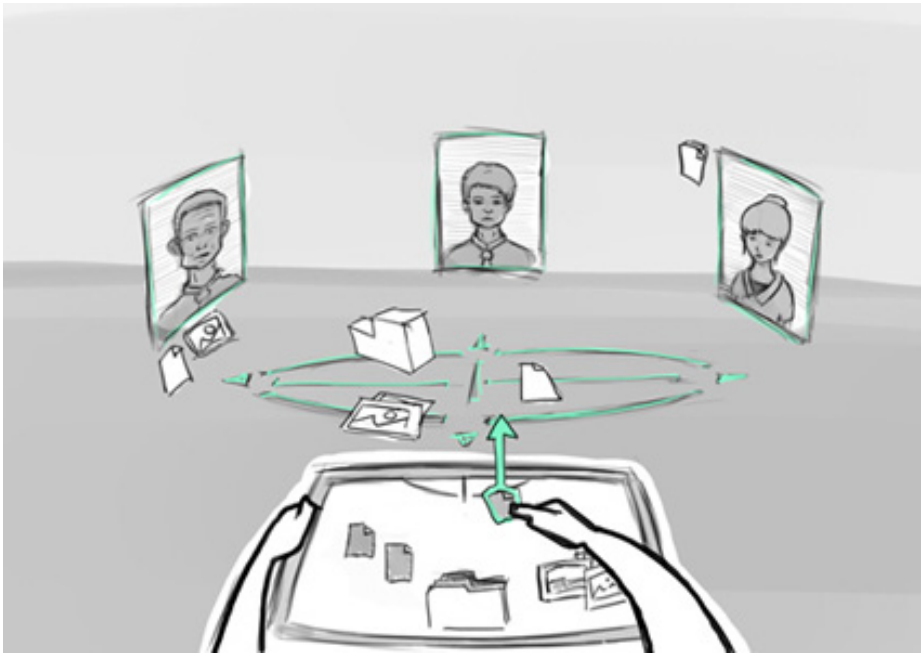


Figure 18: A concept sketch of how an implementation could look like

3.1.3 Inspiration

During the design process the prototypes explored were inspired by science fiction movies and animations that show fictitious conferencing systems. For example, in the Japanese animation series *Neon Genesis Evangeleon* there was a conferencing system consisting of a single user surrounded by tall dark obelisks, illustrating the members of the conference, as shown in Figure 19. When viewing the animation it becomes clear that the system uses spatial audio as well as visual feedback to help identify who is talking. The whole conference seems to be virtual, except for main user's desk. While the system does not look very practical in the animation, it speaks to the imagination and adds to the mysterious nature of the conference.

Another animation that has inspired me is *Summer Wars*. In this movie, the internet and all applications are a visual world called Oz (see Figure 20). People can travel through Oz and talk or work together with others through their avatar. The colourful world is most likely used as a metaphor of the internet and the complex network applications that run through it, rather than a real digital 3D world.

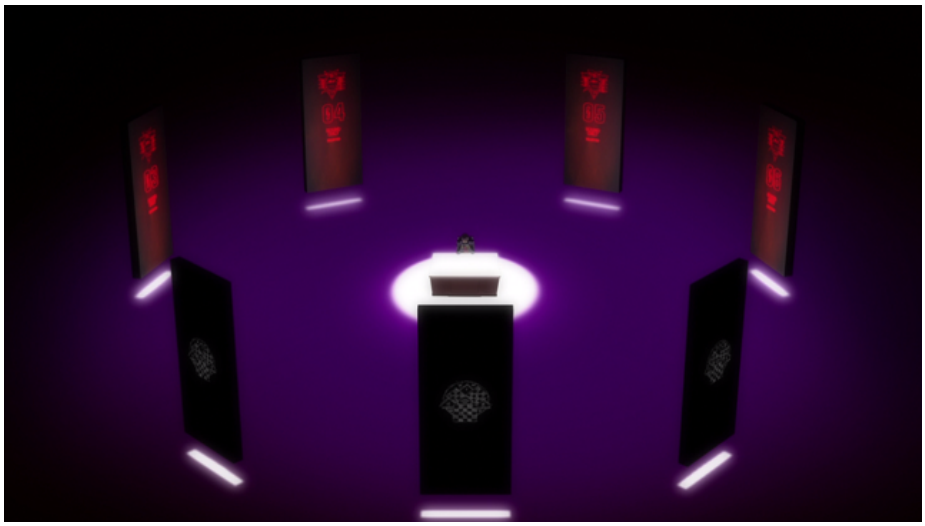


Figure 19: A fictitious conferencing system used in Neon Genesis Evangeleon

Figure 20: The Oz hub, full of travelling avatars

A very well known science-fiction remote conferencing system is shown in *Star Wars*. In these movies remote speakers appear as a blue see-through hologram in the meeting room (Figure 21). Other than the blue tint, remote users appear as if they are physically in the same location as the rest of the participants.

Figure 21: A holographic participant in the conference

Of course, in a practical application there is generally no desire for mystery. In contrary, the desire is to create applications that provide a user with information that is easier to understand than existing alternatives.

3.2 Final Concept and Interaction

Based on the brainstorming conducted and inspired by science fiction and other sources, the final concept developed was an Augmented Reality Conferencing room, in which the main user plays a central role. The user will be able to initiate video calls with multiple people, and has control over the location of these people in 3D space, moving them around at will. For instance, a user can decide to group two people to the right based on their field of expertise or maybe their physical location. Another person could be placed directly in front of the user, for instance if this is the main contact the users needs a conversation with at the moment. A head orientation tracker will allow the user to look around using head movements or by rotating the whole body, creating a body-stabilized Augmented Reality environment. The system also uses spatial audio to strengthen the feeling of presence and to add depth to the virtual space. In this case, each person in the call will act as an audio source and the user will perceive their voices from the correct direction and distance. A simple concept sketch is shown in Figure 22.

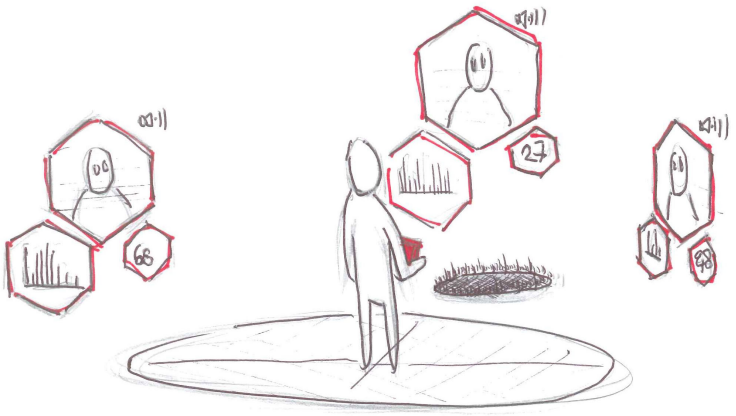


Figure 22: A sketch of an AR conference room

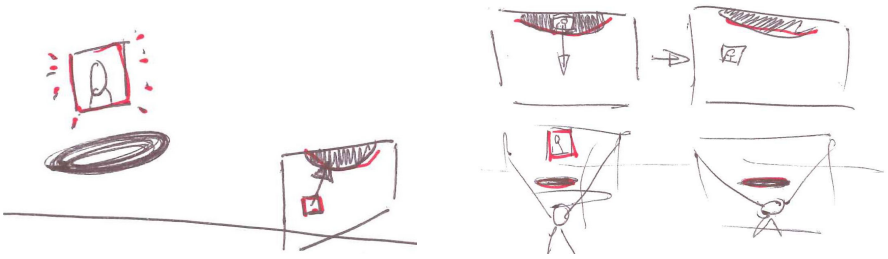
The other part of the system is a file sharing system. The user will be able to use HHD and head movement to share files with the people added to the conversation. The file sharing aspect of this proposed system opens interesting opportunities for interaction design, so the research was focussed on this functionality. There are two interaction methods explored; an *egocentric* and an *exocentric* method.

In the egocentric method, actions on the tablet interface generate results depending on where the user is looking in the augmented reality view. This means that when sending a file, it will arrive where the user is looking. To send the file to a certain virtual participant, the user must look at that person and swipe the file from the tablet up to the HMD. When sharing a file publicly, the user can swipe it down as a metaphor of dropping a file on the table or floor where everybody can see it.

For the exocentric method, the head orientation and interactions on the tablet are independent from each other. A user does not have to look at a person, but they can drag items directly on the faces of the conference participants shown on the HHD. The tablet shows a 'radar' view that updates the display with the users head-rotation.

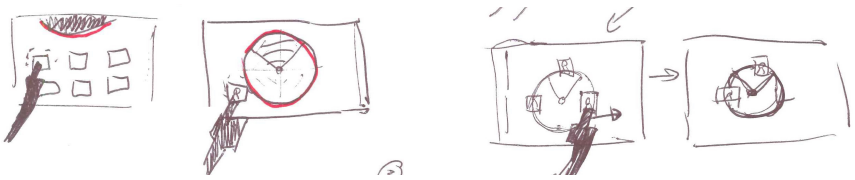
Early sketches of these interfaces are shown in Figure 23 for the egocentric condition and Figure 24 for the exocentric condition. Figure 23 A shows how a contact is dragged to a designated area on the HHD. This area corresponds

with a cursor disc shown on the HMD. When the dragged into that direction, it appears to pop out of the cursor disc. Sketch B shows how a person is removed from the AR conference space. The user moves the cursor disc to the contact shown on the HMD. This selects the contact and shows it on the HHD. The user then drags the contact out of the designated area onto another area of the HHD. Figure 24 A shows how to add a person using the exocentric interface. A user starts a drag action by holding their finger on a contact. The radar disc then appears, and a user can drop the contact anywhere on the radar to place them in a corresponding area on the HMD. A contact is easily removed from the conference by dragging the contact from the radar view and dropping it somewhere else on the HHD. Another iteration of sketches is shown in Figure 25 for the egocentric interface and in Figure 26 for the exocentric. These are more detailed and explore slightly different interactions. These interactions were later applied to sharing files.



(a) Adding people (b) Removing people

Figure 23: Early sketches for the egocentric condition



(a) Adding people (b) Removing people

Figure 24: Early sketch for the exocentric condition

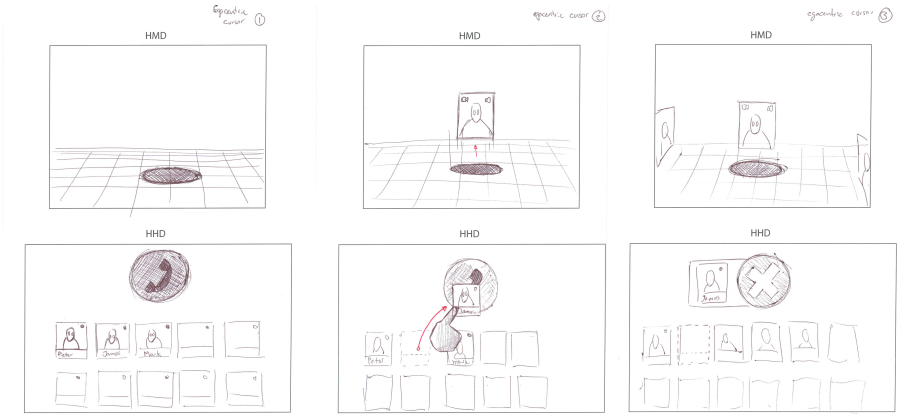


Figure 25: Egocentric HMD and HHD view. From left to right: Idle view, adding person, removing person.

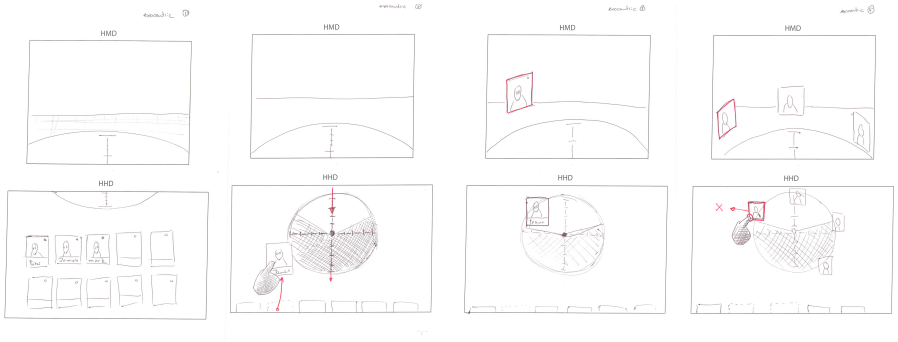


Figure 26: Exocentric HMD and HHD view. From left to right: Idle view, adding person, Idle view with person, removing person.

3.3 Summary

In this chapter we have described the Interaction Design process followed in our research, and in particular how we arrived at the design of our prototype interface. As mentioned, in designing the system, our intended user is someone who needs to perform a remote maintenance task, or similar remote consultation involving document sharing. Based on this user's needs we had a brainstorming session to explore different ways that an AR conference tool could be created using a HHD and HMD. From the ideas collected and other sources, we sketched out a prototype system that uses a spatial metaphor to create the AR conferencing space. Virtual images of remote people and spatial audio are used

to identify who the participants are and when they are speaking. Ego-centric and exo-centric methods for document exchanged were also explored. In the next chapter we describe how we developed a working prototype to evaluate the different interaction methods proposed in the design.

4 Implementation and Prototype

After using a brainstorming process to arrive at a design concept, a working prototype interface was created. In this chapter we describe the hardware and software developed to build the prototype. The overall goal was to create a working system robust enough for user evaluation. In this chapter we describe the final system in detail, while in the next chapter we present a user evaluation conducted with the system.

4.1 The Hardware

Our interface design focuses on showing an AR conferencing view in a Head Mounted Display and using a hand held display to interact with the content. In terms of the hardware, the most important items in the project will be the HMD, the head tracker used to measure the user's viewpoint, and the hand held display used for interaction. Several different hardware devices will be used in this project and this section reviews the available options that were considered and describes the final selected hardware.

4.1.1 Available Technology

There are many different devices available that could be used in this project for both the consumer market and the research market. The HITLabNZ own several different Head Mounted Displays, head trackers and Android tablets.

Head Mounted Display

One of the head-mounted displays available is the Vuzix Wrap 1200VR (Figure 27) which is a stereo video display with LCD panels that supports a relatively high resolution of 1280 by 720 pixels. This makes the display sharp, and the overall design is in a sunglasses form factor and so is comfortable to look at. The glasses are not see-through and block most of the light coming in from the outside world. They also support an optional orientation tracker that can be used to find the current orientation of the head. The glasses are connected to a computer using standard VGA and USB cables.

However, there are some disadvantages with the 1200VR. The field of view only spans about 35 degrees, so they provide a less immersive experience. The device is also slightly unbalanced which makes it uncomfortable to wear for an extended period of time and also causes it to slip off the nose easily. The build quality is also mediocre; they fall apart easily and may need to be reassembled.

Figure 27: The Vuzix Wrap 1200VR Head Mounted Display

Despite this, these glasses are still a good option because of their availability and their orientation tracker. Since the Vuzix 1200VR is not an optical see-through device it would either be necessary to add a webcam to the glasses and make them into a video see-through device, or cut off one of the eye pieces, so the user can see the real world with their natural eyes.

Another available HMD is the Brother AirScouter (see Figure 28). This is a monocular, optical see-through device with an excellent display quality and brightness. The resolution is 800 by 600 pixels, which is good compared to the size of the device. The device is very light and worn on a set of regular glasses that keep it in place. Even though the display is see-through, under normal conditions the background does not interfere with the displayed image. The monocular display can be attached to either side of the glasses, which makes it easy to use for both people with left eye-dominance and right eye-dominance. The display size is that of a 16 inch display at a distance of approximately 1 meter, spanning about 22.4 degrees.

The installation of this device is slightly more cumbersome than the Vuzix glasses, since instead of a VGA or other standard display connector it uses two USB connectors. The drivers that are needed for the device cause some problems on computers with an Intel graphics card, and can only be used with NVidia or AMD graphics cards. This display also does not provide a built-in orientation tracker, so it would need a separate device for tracking.

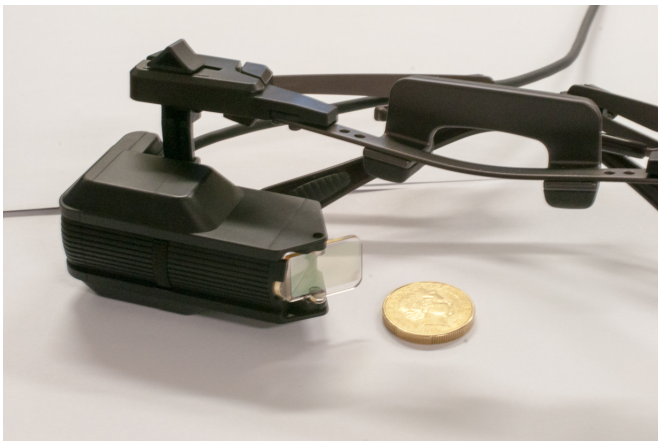


Figure 28: The Brother AirScouter Heads-up Display

Head Tracker

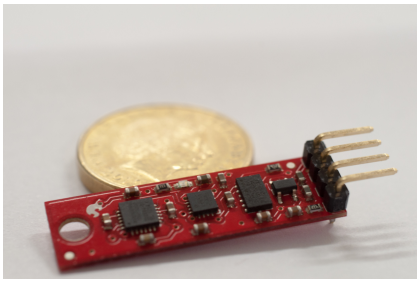
There are many trackers available that will track the orientation of the head. Some trackers can be external and work with camera's or infra red light, however, for this project it is a necessity to have a tracker that is attached to the head since portability and manoeuvrability is very important.

As noted earlier, the Vuzix glasses can be outfitted with an optional head tracker that works well but needs proprietary drivers and code. This makes it hard if not impossible to use the Vuzix tracker with a different set of glasses. Another available sensor is the InterSense iTrax (model 100-ITRAX-0002), see Figure 29. Although this sensor is slightly bigger than the Vuzix sensor, it is usable with any HMD and has a reset button that allows the tracker to be set to its zero point at the start of every use. This function can be very convenient.

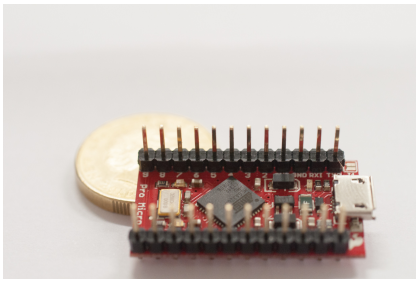
A third option is to use an Arduino board, such as an Arduino Micro, together with a 9 Degrees of Freedom tracker (SEN-10724) and create a custom tracking sensor (see Figure 30). This board consists of an accelerometer, magnetometer and gyroscope. While some of the sensors on the board can sample at 3200Hz, the drivers available for this board and USB limitations allow for updates at a maximum pf 50Hz. Implementing this tracker will be a little more effort in the short term, but might help a lot in the future since it is completely independent from other hardware and also easily replaceable. This hardware configuration is also smaller than the InterSense tracker and widely supported by the Arduino community. A tracker like this would send data to a computer by using a serial connection, which means that it is relatively easy to use, and

Figure 29: The intersense iTrax orientation tracker

no proprietary software or libraries are required.



(a) A 9 Degrees of Freedom tracker



(b) The SparkFun Pro Micro

Figure 30: Components for a head tracker

4.1.2 Final System

The final system used the Brother AirScouter as the head-mounted display, because it was light-weight, had a good display quality and was see-through. A 9 Degrees-of-Freedom orientation sensor (SEN-10724) was attached to the display. The sensor was connected to a SparkFun Pro-Micro Arduino compatible micro controller with an ATMega 32U4 chip. This sensor uses a magnetometer, accelerometer and gyroscope to track the orientation of the sensor. The Arduino board and sensor board were both encased in a 3D printed housing for protection and simplifying the mounting of the sensor to the heads-up display, see Figure 31. Users also wore a pair of Plantronics GameCom 377 specialized headphones that could simulate surround-sound. For the hand-held display an Asus TF700 Android tablet was used. This tablet has a 10.1 inch screen with a resolution of 1920 by 1200 pixels and runs on a 1.60GHz NVIDIA Tegra 3 T33 processor. It has 1GB of RAM and 32GB of storage (Figure 32).

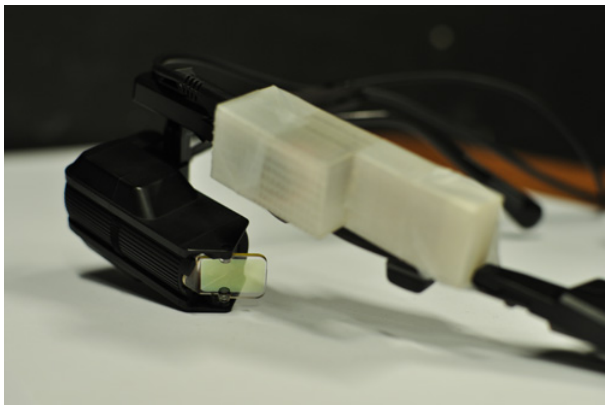


Figure 31: The Brother AirScouter with attached orientation tracker

(a) Headphones

(b) Asus TF700 tablet

Figure 32: Headphones and tablet

4.2 The Software

The prototype software consisted of two parts. One was built for Windows and displayed the AR scene on the HMD (the AR Application). The other part consisted of an interface on the Android tablet that allowed for interaction with the world viewed on the HMD (the Hand held Application). The AR application is build using openFrameworks [22], which is an open source C++ collection of libraries created to simplify the creative process in software design. It combines commonly used software such as OpenGL, GLUT, Quicktime, OpenCV in a convenient and user-friendly wrapper.

Following the concept design in the previous chapter, the AR application should show a virtual conferencing space with 3D representations of the different people in the conference, and spatial audio for their conversation. In addition, on the hand held device there should be an application that allows the user share

documents and information between the participants. In order to support this design, the AR and tablet applications have to have the following properties:

AR Application

- showing 3D graphics scene
- using head tracking data to set the user's viewpoint
- overlaying graphics on the real world to create AR view
- providing spatial audio feedback
- networking support to the tablet application for data exchange

Hand held Application

- showing 2D graphics representing the conference layout
- showing 2D graphics representing the data to be send to conference participants
- supporting 2D touch input
- networking support to the AR application for data exchange

Figure 33 shows a high level architecture of the various system components that need to be developed for the prototype. In the remainder of this section we describe these software components in more detail.

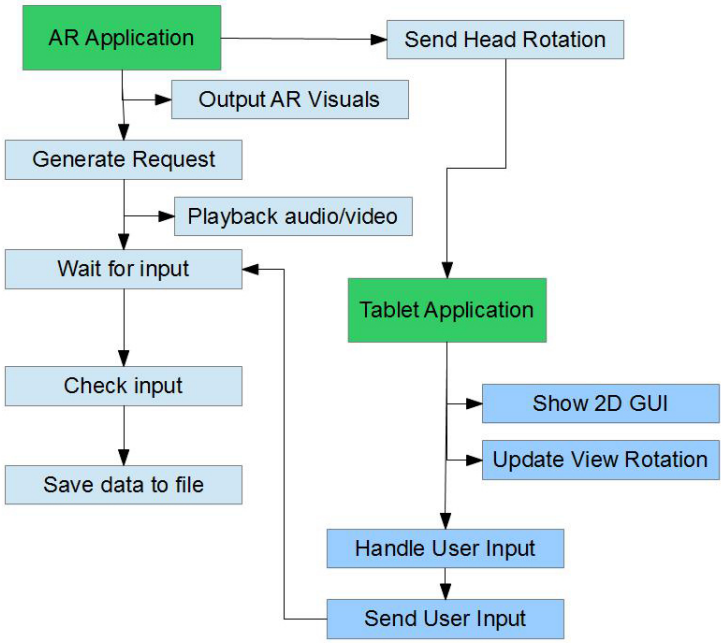


Figure 33: High level block diagram showing the various software components and how they relate to each other

4.2.1 The AR Application

To develop the AR application, openFrameworks was used with the ofxOsc and ofxXmlSettings addon libraries. OpenFrameworks was chosen because it could be used to provide the 3D graphics, networking and audio support needed by the AR application. To enable networking between the AR application and the handheld tablet, the ofxOsc library is used which is an addon that provides OSC network communication functionality. OSC (Open Sound Control) is a networking protocol that sits on top of the UDP data transfer protocol, and provides the ability to receive and send multiple parts of information easily and with very low latency [21]. The advantage of OSC compared to bare UDP is that the content of messages can be more formatted and descriptive. In the prototype application, the program looks for OSC messages that start with the Address *Object*. As soon as a message like this is received, it will go through the list of arguments that were sent with the message. The following is an example of how a message is sent with ofxOSC.

```
ofxOscSender sender;      //create sender object
sender.setup(HOST, PORT_OUT); //set host and port
```

```
ofxOscMessage m;      //create message object
m.setAddress("rotation"); //set an address
m.addFloatArg(yaw);       //add a value

sender.sendMessage(m);      //send the message
```

To play back sound files and generate spatial audio in the interface the FMOD sound library was used. [30]. An early prototype based on the FMOD sound library is shown in Figure 34. This implementation shows a top-down view of a virtual conference room, in which every dot is either a listener or a sound origin. A listener is in this case a virtual head with two virtual microphones as ears, that are used to model the spatial sound. A listener has a position and facing direction, while sound origins have a position and travelling direction. The positions of the listener and sound origins could be dragged around with the mouse. When wearing a pair of headphones the sounds clearly changed when dragging them around on the screen. The effect was especially clear when dragging with a circular movement around the listener. While this prototype was just a simple 2D representation of a listener and a few sound origins, it had the basic functionality of adding and removing sounds dynamically. This worked well as a proof of concept for the spatial sound.

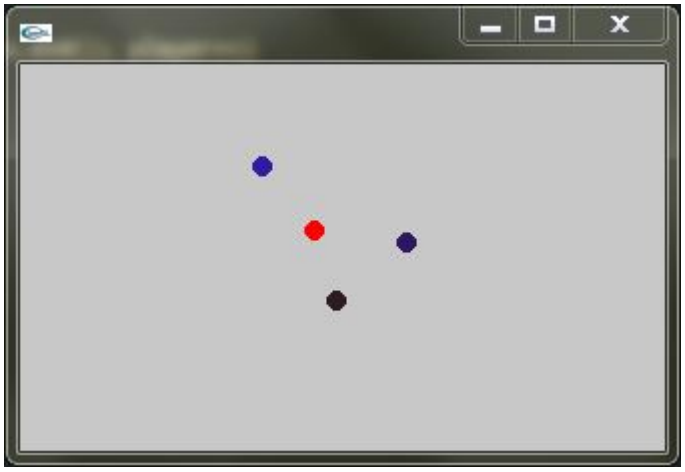


Figure 34: A top down view of a virtual room with spatial audio. The red dot is a listener, while the others are sound origins

To show the AR scene a 3D virtual environment was created, having 3D objects as sound origins and a listener who's head orientation was directly controlled by head movements of the real user. A simple 3D environment was created, in which the listener is bound to the camera and sound origins are

represented by boxes located in the environment,see Figure 35. In this initial version of the prototype, a user had the ability to add, relocate and remove the boxes and the InterSense iTrax tracker was used for head tracking. All three axes of rotation were used to simulate a natural view.

This application was shown to a few people to ask for early feedback on the software. While doing this, it was found that every individual does not easily recognize simulated spatial audio. Some people can very precisely indicate where a sound is coming from, while others said that they did not hear any differences. We assumed that when spatial audio would be combined with head-tracked visuals, this would enhance the perceived spatiality significantly more than when used on its own.

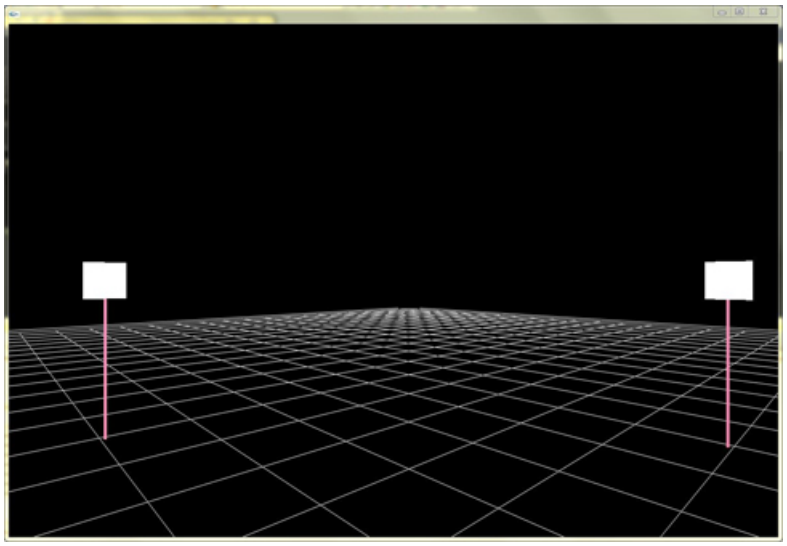


Figure 35: An early build of the 3D environment with the sound origins represented as boxes.

After developing an initial 3D prototype video portraits were added instead of boxes as sound origins (see Figure 36). These video portraits represented remote users in the conference, and could be moved around to change the spatial arrangement of the conferencing space. A red cursor disc was added to select the portraits with and move them around. A user would use natural head movements to look around in the scene and control the cursor disc. The cursor disc appeared to be on the floor, two meters in front of the user. By looking at a portrait, the cursor disc is moved underneath it. When the cursor disc would be underneath one of the video portraits, it would turn green and indicate that the portrait could now be selected. Selection was done by the X key on the keyboard,

when the same key was hit again, the portrait would be deselected. After a portrait was selected, the user could use head movement again to move the cursor disc and portrait, and place it somewhere else in the scene by deselecting it on the desired location. The red lines were used for debugging and were later removed. Because the HMD used was and optical see-through device, all areas that are left black in the application appeared as transparent on the HMD.

In addition to supporting spatial arrangement of conference participants, software was also added to allow file sharing between participants. An OSC receiver was created to support object sharing between the AR and tablet applications. It was built so it could receive messages with the address *"object"* only. A simple test application was written in C++ using openFrameworks to send the OSC messages with the same address. These messages contained a shape (square, triangle or circle) and a colour (red, yellow or blue) and a target, representing the file to be shared. The target was a number that represents the corresponding video portrait and sound origin. When the OSC message was sent, a 3D representation of the shape appeared underneath the target as seen in Figure 36.

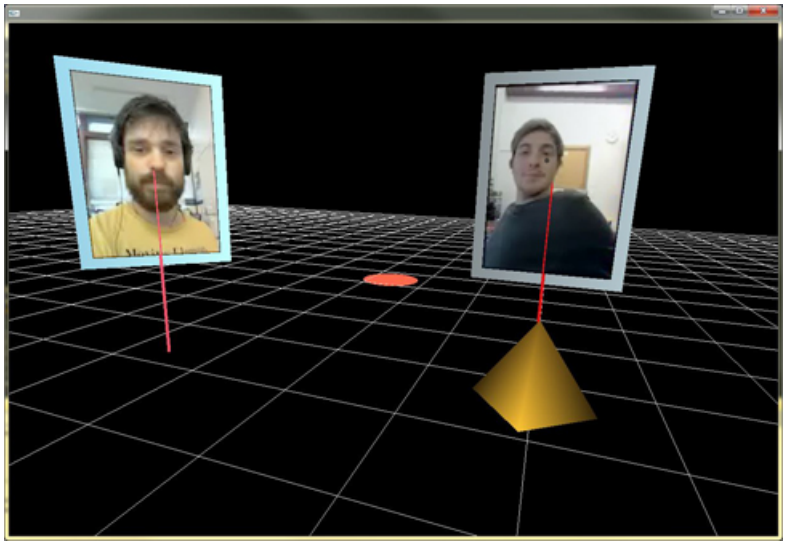


Figure 36: Boxes are replaced by portraits. The program just received the Yellow Triangle message and is showing the 3D representation.

The final version of the AR software had some minor changes implemented that made the scene look more appropriate on the HMD. As it turned out, the portraits were relatively small, so they were moved closer to the user. More

speakers were added in the conference for a total of five virtual people. The user evaluation was going to be performed for a simulated conference in order to control the conference content. So, video and audio samples were recorded for all of the simulated participants. These sample clips simulated requests for the user to send a file to a particular conference participant. There was also a green disc added that appeared to be on the floor 37.

Figure 37: A screenshot of the final AR application. The yellow circle is a visual cue.

4.2.2 The Handheld Application

The hand held Android tablet application was relatively simple, since it only needs to display a 2D user interface, and to send and receive data over a local network to the AR application. Two different user interfaces were created, deliberately designed to be very simple and minimal, with only the most necessary elements. The first was an egocentric view that only showed the files that could be dragged around and an upper and lower area. The second was an exocentric view that showed the files, a radar view, the locations of the virtual people and the view cone of the user.

Both interfaces received some data from the AR application, such as the requested shape, colour and person requesting. This was needed to show visual cues on the tablet. The exocentric view also received the rotation data to update its view accordingly. Both interfaces also had their own way of triggering a sound retry. A sound retry could be triggered by users in the evaluation sessions to play an audio cue again, in case they did not understand it or could not locate it the first time.

The egocentric interface consisted of two drop areas and a centre area with the file representations. The basic metaphor is that users can flick files to the participants in the AR conference. To do this a user first selects a file by holding their finger on it. Then they can throw the file into the HMD view by then swiping it up into the large green area. The file is then sent to the selected conference participant. These are selected by looking at them through the HMD. Alternatively, by dragging and dropping the file into the lower darker green area, a user shares it with everybody (see Figure 38). A sound cue retry was triggered by tapping the large drop area on the top of the screen.

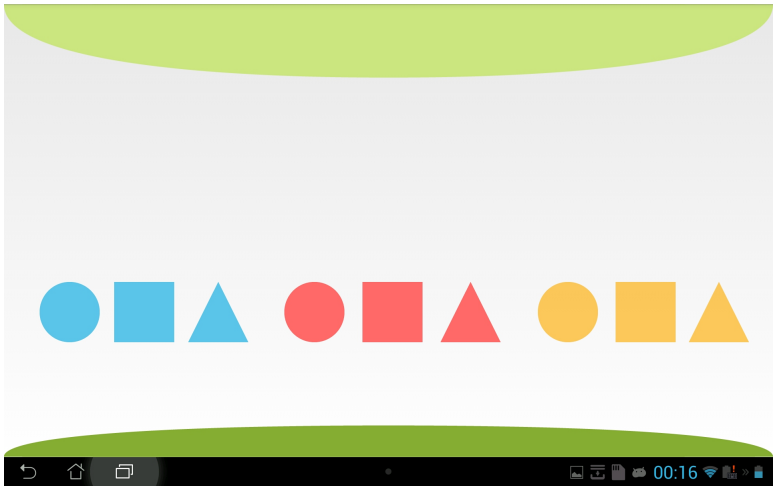


Figure 38: The Egocentric tablet interface.

The exocentric interface showed a radar view of the 3D space shown on the HMD, and the file representations (see Figure 39). The radar view showed all the avatars of the virtual people currently in the conference. The centre area of the radar circle represented the user and the cone in the top part of the circle represents their field of view. When the user rotates their head, the location of the avatars also updated to represent the new orientation. This happened without any noticeable delay. Using touch input, the user could drag the files either onto the avatar of the person they want to send it to, or drop it in the centre of the radar to share a file with everybody. A sound cue retry was triggered by tapping the centre of the radar.

Figure 39: The Exocentric tablet interface.

In order to implement networking between the handheld and AR applications, the JavaOSC library was used for communication [25]. This was because Android by itself does not have an implementation of OSC communication. A simple implementation for receiving an OSC message using JavaOSC is as follows:

```
//create receiver and listen on a port
receiver = new OSCPortIn(portin);
OSCListener object_listener = new OSCListener() {

  @Override
  public void acceptMessage(Date time, OSCMessage message) {
    receive_object(message);
  }
};

//create listener for messages with address "object"
receiver.addListener("object", object_listener);
//start listening
receiver.startListening();
```

4.3 Summary

This chapter described how the prototype was implemented. It reviewed the different options for the hardware and showed what products were used to create the final system. It also went over the software part of the application and how the AR application on the Windows platform works together with the hand

held application on the Android platform. The AR application was created using the C++ language and openFrameworks. It shows five virtual conference participants that could all be able to request a user to send them files. For the hand held application, created in Java for Android, we chose to design a very minimal and clean looking interface that would show the users only what they need. The applications could communicate using the OSC communication protocol.

5 User Evaluation

This chapter describes a comparative study done between the different interfaces and cuing methods that were designed. The goal of the experiment is described first, after which the study design is explained. The chapter ends with the results and their analysis. The information and consent form used in the experiment are found in Appendix A, and the subject questionnaire used can be found in Appendix B.

5.1 Evaluation Goal

The main goal of the evaluation was to compare the proposed exo- and egocentric interfaces and the audio and visual cuing methods. The results of this study will be used to determine a set of design guidelines for future applications that use both an HMD and HHD. To achieve this, a user experiment was conducted to compare each combination of interface and cuing method. Measurements include both quantitative and qualitative measures which aim to find both performance statistics as well as subjects' personal responses.

5.2 Design

5.2.1 Hypothesis

The two main hypotheses of the user experiment were:

- H1: There is no significant difference between the exocentric and egocentric interface for sending content from an HHD to an AR conference space on a monocular HMD.
- H2: There is no significant difference between the audio and visual cuing method for sending content from an HHD to an AR conference space on a monocular HMD.

The difference between Egocentric and Exocentric interfaces on an HHD-HMD AR conferencing system have not been explored before. Both interfaces are likely to have their own strengths and weaknesses, and there is no reason to expect a difference in performance between the two. This lead to the formulation of H1.

In the related work section we show that earlier researchers had found spatial audio could be useful for distinguishing speakers in an AR conferencing application, particularly when people are speaking at the same time. However

there has been little research on how visual and audio cues can be used to support document sharing in AR conferencing. Since it is unlikely that people will be requesting files at the same time, in this case audio and visual cues should probably perform equally well. This assumption lead to the formulation of H2.

5.2.2 Materials

The experiment was done in a small 3.5 by 6 meter room. The following materials were used:

- A Brother AirScouter HMD
- The frame of cinema standard 3D glasses to support the HMD
- A Plantronics GameCom 377 audio headset
- An Asus TF700 tablet
- A desktop computer
- A wifi adapter for the desktop computer
- A laptop computer
- Two chairs
- Two tables
- One Arduino based headtracker in a 3D printed case

One of the tables was used by the experimenter. On this table the desktop computer and other materials were placed. A second table was used by the subject during the reading and signing of the information and consent forms, and when filling out the online questionnaire on the laptop. However, during the actual experimental tasks the user was required to stand up away from the chair and table.

The AirScouter display was connected to the desktop computer by two USB cables, one to supply power and the other to supply the display feed. The headtracker that was attached to the AirScouter had a third USB cable running to the computer. A fourth USB cable was used to connect the headphones. All cables were bound together with tape, to form one single cable and avoid clutter and confusion.

5.2.3 Procedures

To evaluate the different interfaces and interaction methods of the application, a within-subject study was run with four different conditions. There were two independent variables, one being the cue method, the other being the type of the tablet interface. To study the effects they have in different combinations, the following four conditions were chosen:

1. Spatial Audio Cue & Egocentric interface. In this condition files are requested with the use of spatial audio cues, and the subject uses the egocentric HHD interface.

2. Spatial Audio Cue & Exocentric interface. In this condition files are requested with the use of spatial audio cues, and the subject uses the exocentric HHD interface.

3. Visual Cue & Egocentric interface. In this condition files are requested with the use of visual cues, and the subject uses the egocentric HHD interface.

4. Visual Audio Cue & Exocentric interface. In this condition files are requested with the use of visual cues, and the subject uses the exocentric HHD interface.

The order of the conditions varied per subject using a Latin square to counterbalance learning effects. At the start of the study the user was first asked to read the information sheet and read and sign the consent form. They were then asked to fill in some general questions in an online questionnaire. This questionnaire recorded data such as age, gender and previous experience with VR, AR and tablet interfaces. Afterwards, the subject was told exactly what the research was about, and showed the interfaces and explained the cuing methods.

After the system was explained the researcher helped the subject with putting on the equipment. If the subject did not wear glasses, the frame of the cinema 3D glasses was used to support the HMD. After the HMD was in place, the headphones were put on. The subject was then given the tablet. A person wearing the full system is shown in Figure 40.

The experiment explored the ability of subjects to be able to recognize when participants in the AR conference were requesting a file and being able to deliver the file successfully to the participant. The requests were performed by the

Figure 40: User wearing the system

simulated virtual conference participants using the two cuing methods. Subjects would then deliver the files with the two HHD interfaces.

For each condition, a subject was given five training tasks, during which time all their questions about the task were answered. After this the subject was given 20 experimental tasks. The results of these 20 tasks were recorded by the software and saved as an XML document. The following variables were recorded.

- Time taken per task
- Time taken to complete all 20 tasks
- Requested file
- Sent file
- If the sent item was correct
- Number of retries (audio only)
- Total amount of Yaw, Pitch and Roll the user's head made to complete a task

After the total of 25 tasks for each condition, the user was asked to fill in a questionnaire focussing on usability and immersion. When the subject finished the questionnaire the next condition would be started.

5.2.4 Measurements

Several different quantitative measures were recorded during the study. Some of them were recorded by the application itself, while others were recorded using a questionnaire. The questionnaire also recorded qualitative measures.

Session performance time was measured by the application and recorded the time it took for a subject to complete a set of 20 tasks using one of the four conditions. This was measured in seconds from the moment the session was started by the researcher until the 20th task was completed.

The task results were also measured by the application. The task results record whether a task was completed successfully or whether they user made a mistake. Whether a subject completed the ask successfully or not was determined by comparing the requested file and the requester with the sent file and the target. If one or more did not match, the task was not completed. successfully.

For the conditions that made use of the sound cue, the application also recorded the amount of retries the subjects activated. A retry was activated by the subject when the sound cue is not fully understood or the subject cannot find the file requester in the 3D AR space. When a retry was triggered, the sound and video play once more. The study subject could use this as much as needed, however, every retry was recorded.

The final measure recorded by the application was the head movement. The head movement was recorded in degrees, per task in all three axes. This data was recorded by taking the values of the orientation sensor and subtracting the previous value, before taking the absolute value. This value was then added to the total amount of movement.

```
total_yaw += abs(yaw - prev_yaw);
```

The questionnaire recorded some basic demographic data of the subject, as well as quantitative and quantitative data about what the subjects thought of the conditions. The questions dealt with the feelings of engagement, ease of use, ease of learning and fun. Most questions utilized a 5-point Likert scale. The questions that were asked are shown in the results section later in this chapter.

5.2.5 Subject Participants

In the experiment, 2 subjects were selected as part of an initial pilot study, the results of these subjects do not count towards the final results of the study, and were exclusively used to tweak the system and find bugs before the real trials started. For the real study, 17 people were asked to participate. Eventually, only the data of 16 of these was used. Due to a high assumption of bias the results of one of the subjects were excluded from the measurements. This user was very familiar with all voices used in the environment, making the tasks extremely easy. Of the remaining subjects a few were students from the HIT Lab NZ itself, while most were recruited from one of the student clubs at the University of Canterbury.

Of the 16 subjects, 12 (75%) were male and 4 (25%) were female. Most subjects were students at the University of Canterbury and between 20 and 30 years old. Not all subjects were native English speakers, but all had a good level of understanding and speaking. The only prerequisites for subjects to participate were normal or better hearing and normal (colour-) vision. Most subjects did not have a lot or any experience in VR and AR, but since they volunteered as a subjects for this research, the assumption can be made that they are interested in AR and interface technology.

5.3 Results

The results section is divided in three main parts. The first section will evaluate the data that the application has recorded, while the other two parts evaluate the data recorded by the questionnaire. The questionnaire partly recorded quantitative data with 19 questions per condition using a Likert scale and 3 rated questions, and partly it recorded qualitative data with open questions where subjects could write down their thoughts on the system.

5.3.1 Quantitative Measures: Application

The Quantitative measures recorded by the application included the average performance time in (in seconds), the task results (true or false), amount of retries (only used for audio cues) and total amount of head movement (in three axis, in degrees, measured per task).

Since the study design was within-subjects and there were two factors with two levels each, a two-way repeated measures ANOVA test was used to test for statistically significant interaction between the factors. If statistical significance

was found between conditions, a pairwise Bon Ferroni test was done to determine the exact location of the interaction.

Session Performance Time

Table 1 shows the average time for subjects to complete all 20 tasks in each condition. A significant difference in performance time was found between the different conditions. A repeated measures two-way ANOVA test with a Greenhouse-Geisser correction determined that the main effect of both the interface as cuing method on the session time was statistically significant. Using the Exocentric interface decreased the session time significantly compared to using the Egocentric interface ($F(1,15) = 43.869$, $p < 0.001$). The Visual cue method also significantly decreased the session time compared to the Sound cue method ($F(1,15) = 47.418$, $p < 0.001$).

The test also showed that there is a statistically significant interaction between the cuing method and interface ($F(1,15) = 31.418$. $p < 0.001$), see the graph in Figure 41. The graph shows that overall the session time in the conditions with the Visual cue were the lowest. It also shows that for both cue methods the session time was lower for the Exocentric interface. This effect is especially noticeable for the Visual cue method.

Session Time		
Condition	Mean (s)	Std. Dev
EgoSound	150.1713	31.41379
EgoVisual	136.3972	37.70108
ExoSound	140.6694	27.16970
ExoVisual	56.0028	23.26528

Table 1: Mean session times (in seconds) for different conditions

A Simple Mean Effects ANOVA test with the Bon Ferroni adjustment revealed where the interaction between the factors is taking place. For the Visual Cue method, using the Exocentric interface over the Egocentric interface significantly decreased the session time($F(1,15) = 55.350$. $p < 0.001$). The type of interface used had no significant effect when using the Sound Cue method.

When using the Exocentric interface, using the Visual Cue significantly decreased the session time compared to the Sound Cue method. ($F(1,15) = 103.095$, $p < 0.001$). The type of cue method had no significant effect when

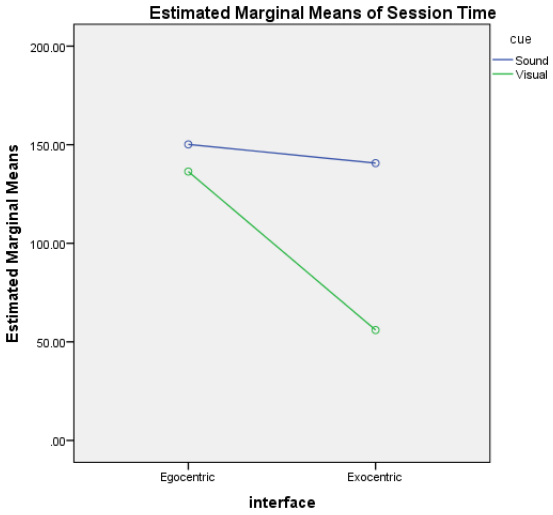


Figure 41: Interaction between the conditions

using the Egocentric interface.

Overall subjects performed the fastest in the condition of Visual Cue + Exocentric.

Task results

Table 2 shows the average percentage of tasks subjects completed successfully for each condition. No significant difference was found in the task results measure. A repeated measures two-way ANOVA test with a Greenhouse-Geisser correction shows that there was no statistically significant difference cause by the interface ($F(1, 15) = 26.898$, $p = .642$) nor by the cue method ($F(1, 15) = 29.754$, $p = .665$).

Task Results		
Condition	Mean (%)	Std. Dev
EgoSound	.8594	07576
EgoVisual	.9344	.05391
ExoSound	.9375	.05323
ExoVisual	1.0000	.00000

Table 2: Mean task results (percent) for different conditions

Retries

No significant difference was found in the retries measure. Retries were initiated by the subject when they require the audio sample to be played more than once. Subjects could initiate a retry as often as needed. Since this measurement was only relevant to the Sound Cue conditions, a one-way ANOVA test was done on the measurements shown in Table 3. There was no statistically significant difference between the two interfaces ($F(1,15) = .288$, $p = .600$).

Retries		
Condition	Mean (ms)	Std. Dev
EgoSound	9.81	4.086
ExoSound	10.38	4.911

Table 3: Mean amount of retries for different conditions

Head Movement

Table 4 shows the total average head movement about each of the three rotational axes in degrees. There was a significant difference in the amount of head movement between the conditions. A repeated measures two-way ANOVA test with a Greenhouse-Geisser correction shows that the head movement differed statistically significantly between the different conditions for all axes.

Head Movement						
	Roll		Pitch		Yaw	
Condition	Mean (ms)	Std. Dev	Mean (ms)	Std. Dev	Mean (ms)	Std. Dev
EgoSound	46.5600	13.73683	54.1148	25.51074	295.6890	99.93012
EgoVisual	44.4140	15.36760	64.8199	30.78711	294.3197	87.78461
ExoSound	41.8632	14.40881	54.7854	21.71702	270.4524	106.78961
ExoVisual	14.7681	15.79864	18.4707	23.82700	54.4277	97.60407

Table 4: Mean head movement per task for different conditions

The interface type significantly changed the total movement on all axes (roll ($F(1,15) = 33.223$, $p < 0.001$), pitch ($F(1,15) = 9.804$, $p < 0.05$), yaw ($F(1,15) = 28.076$, $p < 0.001$)).

The cue method also significantly changed the total movement on all axes (roll ($F(1,15) = 20.005$, $p < 0.001$), pitch ($F(1,15) = 18.463$, $p < 0.05$), yaw ($F(1,15) = 25.546$, $p < 0.001$)).

The test also showed a significant interaction between the conditions for all

axes (roll ($F(1,15) = 22.672$, $p < 0.001$), pitch ($F(1,15) = 28.741$, $p < 0.05$), yaw ($F(1,15) = 13.783$, $p < 0.05$)).

In case of the Roll axis (Figure 42) there was a decrease in movement visible for both cue methods when the Exocentric interface was used instead of the Egocentric interface. The change of interface had the greatest impact on the Visual cue method. Overall there was less movement in the Visual cue method for both interfaces. The change of the cue method had the greatest impact for the Exocentric interface. However, the condition of Egocentric + Visual cue had slightly more movement than the Exocentric + Sound cue condition.

For the Pitch axis (Figure 43) there is a slightly different shape noticeable in the graph. It shows a cross-over interaction, in this case suggesting that both the interface and cue method had a reversed effect on each other. However, since the line for the Sound cue is almost horizontal, this can't be said for certain. What can be said, is that while the movement on the pitch axis for the Visual cue was higher than that for the Sound cue when using the Egocentric interface, it was less when the Exocentric view was used.

The graph for Yaw (Figure 44) shows that while there was almost no difference for either of the cue methods when the Egocentric interface was being used, the amount of head movement in the Yaw axis drops a great amount for the Visual cue method when using the Exocentric interface. It also dropped for the Sound cue method, but not nearly as much.

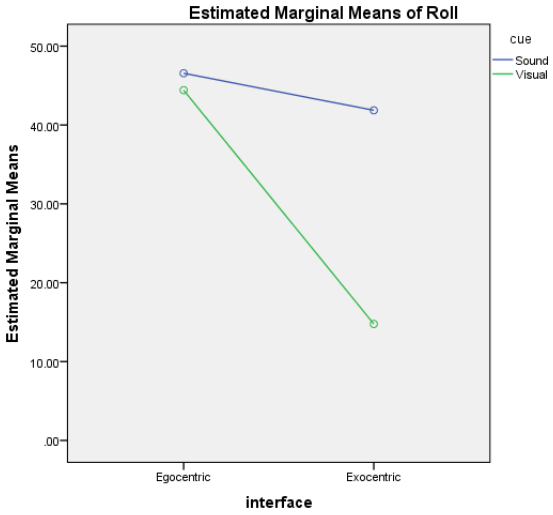


Figure 42: Interaction between the conditions for Roll

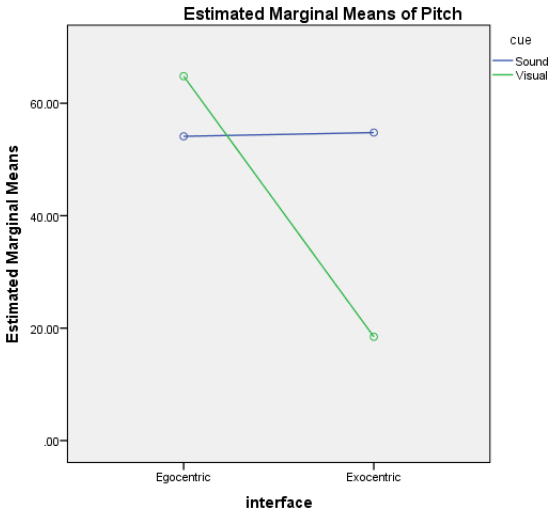


Figure 43: Interaction between the conditions for Pitch

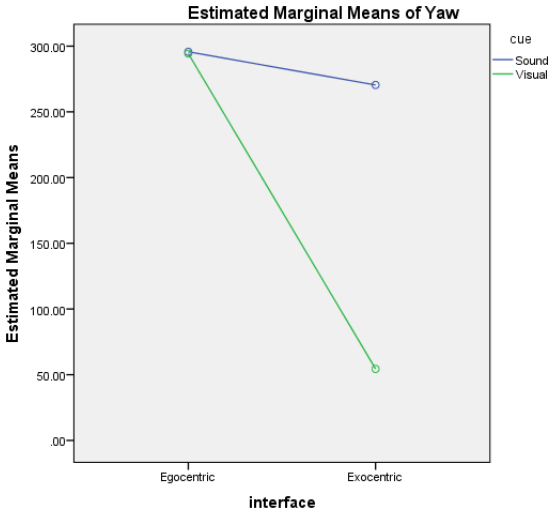


Figure 44: Interaction between the conditions for Yaw

A Simple Mean Effects ANOVA test with the Bon Ferroni adjustment revealed where the interaction between the factors is taking place.

For the Visual Cue method, using the Exocentric interface over the Egocentric interface significantly decreased the head movement in all axes (roll ($F(1,15)= 39.614$ p <0.001), pitch ($F(1,15)= 20.748$ p <0.001), yaw ($F(1,15)= 52.711$ p <0.001)). The type of interface used had no significant effect when using the Sound Cue method.

When using the Exocentric interface, using the Visual Cue method significantly decreased the head movement in all axes compared to the Sound Cue method (roll ($F(1,15)= 37.762$ p <0.001), pitch ($F(1,15)= 33.943$ p <0.001), yaw ($F(1,15)= 33.331$ p <0.001). When using the Egocentric interface, the type of cue method used only had a significant effect on the total movement in the pitch axis and was lower when using the Sound cue method compared to the Visual cue method ($F(1,15)= 6.597$ p <0.05).

Overall a subject moved their head the least in all 3 axes for the Exocentric + Visual Cue conditions and the pitch and roll were highest for the Egocentric + Visual Cue method, while yaw was the highest for the Egocentric + Sound Cue condition.

Summary

The key results found in the recorded data show that the use of the Exocentric + Visual cue condition generally results in significant faster performance and less head movement than other conditions. When using the Visual cue method, the Exocentric interface always results in a significant improvement in session performance time and amount of head movement. In case of the Sound cue method, there are no significant differences between the two interfaces.

5.3.2 Quantitative Measures: Questionnaire

The quantitative part of the questionnaire consisted of 17 questions answered on a Likert scale ranging from 1 (totally disagree) to 5 (totally agree). In a second part subjects where asked to rank the conditions on three different aspects from 1 (best) to 4 (worst). These questions recorded the subjects' thoughts about the conditions. The mean values of the subject responses of the 17 Likert scale questions are shown in table 5.

Mean Likert scale responses									
Condition	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
EgoSound	**3.19**	**3.38**	**3.06**	**3.19**	**4.25**	**3.06**	**3.50**	3.31	**3.31**
EgoVisual	**3.63**	**3.75**	**3.75**	3.69	4.44	**3.19**	**3.44**	3.31	3.31
ExoSound	**3.75**	**3.75**	**3.75**	3.56	**4.50**	**3.50**	3.56	3.69	3.50
ExoVisual	**4.75**	**4.75**	**4.75**	**4.50**	**4.94**	**4.44**	**4.31**	4.13	**4.31**
	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	
EgoSound	3.06	4.00	**3.00**	**2.38**	2.06	**3.87**	2.06	3.44	
EgoVisual	3.25	4.25	**4.19**	**3.75**	2.06	**3.63**	1.94	4.00	
ExoSound	3.69	4.06	**3.13**	**2.75**	2.25	**4.13**	1.94	3.63	
ExoVisual	3.81	4.44	**4.56**	**4.75**	1.56	**2.56**	1.56	4.06	

Table 5: Mean Likert scale responses (between 1 and 5) for the questionnaire. Significant different values are shown in bold.

To determine whether there were significant differences in the subjects' responses to the questionnaire regarding the different conditions, a one-way ANOVA test (Friedman test) was applied to the mean results. A post-hoc analysis using Wilcoxon signed-rank tests was conducted to determine between which pairs of conditions there were significant differences. There are multiple comparisons to be made, so the Bon Ferroni adjustment was applied. The adjustment value was calculated as follows:

$$.05/(\text{number of pairwise comparisons}) = .05/6) = .0083$$

As a result, $p<.0083$ was used for the post-hoc tests. The ANOVA test and post-hoc results are mentioned below:

Q1: Overall, I am satisfied with how easy it is to use the system.

There was a statistically significant difference in the response to Question 1 between the conditions, $\chi^2(3) = 28.728$, $p < 0.001$. A post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .0083$. There was a statistically significant difference in the response to Question 1 between the Exocentric + Visual cue and Egocentric + Sound cue conditions (Z=-3.477, p=.001), the Exocentric + Visual cue and Egocentric + Visual cue conditions (Z=-3.080, p=.002) and the Exocentric + Visual cue and Exocentric + Sound cue conditions (Z=-3.066, p=.002). See Table 6.

Subjects were most satisfied with the Exocentric + Visual cue condition's ease of use compared to any other condition.

Q1 Pairwise tests						
	EgoVis - EgoSnd	ExoSnd - EgoSnd	ExoVis - EgoSnd	ExoSnd - EgoVis	ExoVis - EgoVis	ExoVis - ExoSnd
Z	-1.748	-2.496	-3.477	-.513	-3.080	-3.066
Asymp. Sig.	.080	.013	.001	.608	.002	.002

Table 6: Wilcoxon signed ranks test results

Q2: It was simple to use this system.

There was a statistically significant difference in the response to Question 2 between the conditions, $\chi^2(3) = 23.595$, $p < 0.001$. A post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .0083$. There was a statistically significant difference in the response to Question 2 between the Exocentric + Visual cue and Egocentric + Sound cue conditions (Z=-3.236, p=.001), the Exocentric + Visual cue and Egocentric + Visual cue conditions (Z=-2.859, p=.004) and the Exocentric + Visual cue and Exocentric + Sound cue conditions (Z=-3.066, p=.002). See Table 7.

Subjects found the Exocentric + Visual cue condition more simple than any of the other conditions.

Q3: I could effectively complete the tasks and scenarios using this system.

Q2 Pairwise tests						
	EgoVis - EgoSnd	ExoSnd - EgoSnd	ExoVis - EgoSnd	ExoSnd - EgoVis	ExoVis - EgoVis	ExoVis - ExoSnd
Z	-1.613	-1.103	-3.236	.000	-2.859	-3.066
Asymp. Sig.	.107	.270	.001	1.000	.004	.002

Table 7: Wilcoxon signed ranks test results

There was a statistically significant difference in the response to Question 3 between the conditions, $\chi^2(3) = 25.767$, $p < 0.001$. A post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .0083$. There was a statistically significant difference in the response to Question 3 between the Exocentric + Visual cue and Egocentric + Sound cue conditions (Z=-3.462, p=.001), the Exocentric + Visual cue and Egocentric + Visual cue conditions (Z=-2.944, p=.003) and the Exocentric + Visual cue and Exocentric+ Sound cue conditions (Z=-2.859, p=.004). See Table 8.

Subjects felt that the Exocentric + Visual cue condition was the most effective compared to the other conditions.

Q3 Pairwise tests						
	EgoVis - EgoSnd	ExoSnd - EgoSnd	ExoVis - EgoSnd	ExoSnd - EgoVis	ExoVis - EgoVis	ExoVis - ExoSnd
Z	-2.351	-2.326	-3.462	-.047	-2.944	-2.859
Asymp. Sig.	.019	.020	.001	.963	.003	.004

Table 8: Wilcoxon signed ranks test results

Q4: I felt comfortable using this system.

There was a statistically significant difference in the response to Question 4 between the conditions, $\chi^2(3) = 13.993$, $p < 0.001$. A post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .0083$. There was a statistically significant difference in the response to Question 4 between the Exocentric + Visual cue and Egocentric + Sound cue conditions (Z=-2.841, p=.004). **See Table 9.**

Subjects felt that the Exocentric + Visual cue condition was more comfortable to use than the Egocentric + Sound cue condition. There were no

statistically significant differences between the remaining conditions.

Q4 Pairwise tests						
	EgoVis - EgoSnd	ExoSnd - EgoSnd	ExoVis - EgoSnd	ExoSnd - EgoVis	ExoVis - EgoVis	ExoVis - ExoSnd
Z	-1.565	-1.222	-2.841	-.535	-1.893	-2.240
Asymp. Sig.	.118	.022	.004	.593	.058	.025

Table 9: Wilcoxon signed ranks test results

Q5: It was easy to learn to use this system.

There was a statistically significant difference in the response to Question 5 between the conditions, $\chi^2(3) = 15.324$, $p < 0.05$. A post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .0083$. There was a statistically significant difference in the response to Question 5 between the Exocentric + Visual cue and Egocentric + Sound cue conditions (Z=-2.810, p=.005) and the Exocentric + Visual cue and Exocentric + Sound cue conditions (Z=-2.646, p=.0081). See Table 10.

Subjects felt that the Exocentric + Visual cue condition was easier to learn than the conditions that used the Sound cue method.

Q5 Pairwise tests						
	EgoVis - EgoSnd	ExoSnd - EgoSnd	ExoVis - EgoSnd	ExoSnd - EgoVis	ExoVis - EgoVis	ExoVis - ExoSnd
Z	-.966	-1.414	-2.810	-.447	-2.530	-2.646
Asymp. Sig.	.334	.157	.005	.655	.011	.008

Table 10: Wilcoxon signed ranks test results

Q6: I believe I could become productive quickly using this system.

There was a statistically significant difference in the response to Question 6 between the conditions, $\chi^2(3) = 21.984$, $p < 0.001$. A post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .0083$. There was a statistically significant difference in the response to Question 6 between the Exocentric + Visual cue and Egocentric + Sound cue conditions (Z=-3.477-3.376, p=.001), the Exocentric + Visual cue and Egocentric + Visual cue conditions (Z=-2.869, p=.004)

and the Exocentric + Visual cue and Exocentric + Sound cue conditions (Z=-2.877, p=.004). See Table 11.

Subjects were most satisfied with the Exocentric + Visual cue condition's ease of use compared to any other condition.

Q6 Pairwise tests						
	EgoVis - EgoSnd	ExoSnd - EgoSnd	ExoVis - EgoSnd	ExoSnd - EgoVis	ExoVis - EgoVis	ExoVis - ExoSnd
Z	-.530	-1.941	-3.376	-1.072	-2.869	-2.877
Asymp. Sig.	.596	.052	.001	.284	.004	.004

Table 11: Wilcoxon signed ranks test results

Q7: The interface of this system was pleasant.

There was a statistically significant difference in the response to Question 7 between the conditions, $\chi^2(3) = 12.523$, $p < 0.05$. A post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .0083$. There was a statistically significant difference in the response to Question 7 between the Exocentric + Visual cue and Egocentric + Sound cue conditions (Z=-2.648, p=.0081) and the Exocentric + Visual cue and Egocentric + Visual cue conditions (Z=-2.658, p=.0079). See Table 12.

Subjects found the use of the Exocentric conditions more pleasant than the Egocentric conditions.

Q7 Pairwise tests						
	EgoVis - EgoSnd	ExoSnd - EgoSnd	ExoVis - EgoSnd	ExoSnd - EgoVis	ExoVis - EgoVis	ExoVis - ExoSnd
Z	-.289	-.312	-2.648	-.632	-2.658	-2.521
Asymp. Sig.	.773	.755	.008	.527	.008	.012

Table 12: Wilcoxon signed ranks test results

Q8: I liked using the interface of this system.

There was a statistically significant difference in the response to Question 8 between the conditions, $\chi^2(3) = 9.275$, $p < 0.05$. A post-hoc analysis with

Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at p <.0083. There were no statistically significant differences between the rest of the conditions. See Table 13.

Q8 Pairwise tests						
	EgoVis - EgoSnd	ExoSnd - EgoSnd	ExoVis - EgoSnd	ExoSnd - EgoVis	ExoVis - EgoVis	ExoVis - ExoSnd
Z	-.061	-1.613	-2.176	-1.428	-2.092	-1.461
Asymp. Sig.	.951	.107	.030	.153	.036	.144

Table 13: Wilcoxon signed ranks test results

Q9: Overall, I am satisfied with this system.

There was a statistically significant difference in the response to Question 9 between the conditions, $\chi^2(3) = 15.243$, p <0.05. A post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at p <.0083. There was a statistically significant difference in the response to Question 9 between the Exocentric + Visual cue and Egocentric + Sound cue conditions (Z=-2.863, p=.004). There were no statistically significant differences between the rest of the conditions. See Table 14.

Subjects found the Exocentric + Visual cue condition more satisfactory than the Egocentric + Sound cue condition. Overall the Exocentric + Visual cue condition was rated the most satisfactory

Q9 Pairwise tests						
	EgoVis - EgoSnd	ExoSnd - EgoSnd	ExoVis - EgoSnd	ExoSnd - EgoVis	ExoVis - EgoVis	ExoVis - ExoSnd
Z	.000	-.905	-2.863	-.905	-2.411	-2.124
Asymp. Sig.	1.000	.366	.004	.366	.016	.034

Table 14: Wilcoxon signed ranks test results

Q10: The interactions with the environment seemed natural.

There was no statistically significant difference in the response to Question 10 between the conditions, however, it was approaching significance $\chi^2(3) = 6.620$, $p = 0.085$.

Q11: I was able to anticipate what would happen in response to the actions that I performed.

There was no statistically significant difference in the response to Question 11 between the conditions, however, it was approaching significance $\chi^2(3) = 7.050$, $p = 0.070$.

Q12: I could easily identify the cues. (The icons or sound that appeared)

There was a statistically significant difference in the response to Question 12 between the conditions, $\chi^2(3) = 29.771$, $p < 0.001$. A post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .0083$. There was a statistically significant difference in the response to Question 12 between the Egocentric + Visual cue and Egocentic + Sound cue conditions (Z=-3.000, p=.003), the Exocentric + Visual cue and Egocentric + Sound cue conditions (Z=-3.407, p=.001), and the Exocentric + Visual cue and Exocentric + Sound cue conditions (Z=-3.360, p=.001). See Table 15.

Subjects found it generally easier to identify visual cues than sound cues.

Q12 Pairwise tests						
	EgoVis - EgoSnd	ExoSnd - EgoSnd	ExoVis - EgoSnd	ExoSnd - EgoVis	ExoVis - EgoVis	ExoVis - ExoSnd
Z	-3.000	-.577	-3.407	-2.571	-1.508	-3.360
Asymp. Sig.	.003	.564	.001	.010	.132	.001

Table 15: Wilcoxon signed ranks test results

Q13: I could easily localize the cues. (The icons or sound that appeared)

There was a statistically significant difference in the response to Question 13 between the conditions, $\chi^2(3) = 36.229$, $p < 0.001$. A post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .0083$. There was a statistically significant difference in the response to Question 13 between the Egocentric + Visual cue and Egocentic + Sound cue conditions (Z=-2.885, p=.004), the Exocentric + Visual cue and Egocentric + Sound cue conditions (Z=-3.601,

p<.001), the Exocentric + Visual cue and Egocentric + Visual cue conditions (Z=-3.025, p=.002), and the Exocentric + Visual cue and Exocentric + Sound cue conditions (Z=-3.555, p<.001). See Table 16.

Subjects found it generally easier to localize visual cues than sound cues. The localization of cues was found the easiest for the Exocentric + Visual cue condition.

Q13 Pairwise tests						
	EgoVis - EgoSnd	ExoSnd - EgoSnd	ExoVis - EgoSnd	ExoSnd - EgoVis	ExoVis - EgoVis	ExoVis - ExoSnd
Z	-2.885	-1.222	-3.601	-2.436	-3.025	-3.555
Asymp. Sig.	.004	.222	.000	.015	.002	.000

Table 16: Wilcoxon signed ranks test results

Q14: I felt confused or disorientated at the end of the session.

There was a statistically significant difference in the response to Question 14 between the conditions, $\chi^2(3) = 9.104$, p <0.05. A post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at p <.0083. There were no statistically significant differences between the conditions. See Table 17.

Q14 Pairwise tests						
	EgoVis - EgoSnd	ExoSnd - EgoSnd	ExoVis - EgoSnd	ExoSnd - EgoVis	ExoVis - EgoVis	ExoVis - ExoSnd
Z	.000	-.832	-1.672	-.711	-1.630	-1.865
Asymp. Sig.	1.000	.405	.094	.477	.103	.062

Table 17: Wilcoxon signed ranks test results

Q15: I felt involved in the virtual environment.

There was a statistically significant difference in the response to Question 15 between the conditions, $\chi^2(3) = 25.351$, p <0.001. A post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at p <.0083. There was a statistically significant difference in the response to Question 15 between the Exocentric + Visual cue and Egocentric + Sound cue conditions (Z=-3.270, p=.001), the

Exocentric + Visual cue and Egocentric + Visual cue conditions (Z=-2.675, p=.007) and the Exocentric + Visual cue and Exocentric + Sound cue conditions (Z=-3.219, p=.001). See Table 18.

Subjects generally felt less involvement in the Exocentric + Visual cue condition than in any of the other conditions.

Q15 Pairwise tests						
	EgoVis - EgoSnd	ExoSnd - EgoSnd	ExoVis - EgoSnd	ExoSnd - EgoVis	ExoVis - EgoVis	ExoVis - ExoSnd
Z	-1.027	-2.000	-3.270	-2.126	-2.675	-3.219
Asymp. Sig.	.305	.046	.001	.033	.007	.001

Table 18: Wilcoxon signed ranks test results

Q16: I experienced delay between my actions and the expected outcomes.

There was no statistically significant difference in the response to Question 16 between the conditions, $\chi^2(3) = 3.702$, $p = 0.295$.

Q17: In the end I became proficient in interacting with the virtual environment.

There was no statistically significant difference in the response to Question 17 between the conditions, , however, it was approaching significance $\chi^2(3) = 6.641$, $p = 0.084$.

5.3.3 Quantitative Measures: Ratings

After completing the 17 Likert scale questions, subjects were asked to rank the four conditions in order according to Ease of Use, Fun and Effectiveness. Table 19 shows the mean results of these rankings.

Ratings (1 = best, 4 = worst)			
	Ease of Use	Fun	Effectiveness
Condition	Mean	Mean	Mean
EgoSound	3.81	3.50	3.63
EgoVisual	2.19	2.75	2.50
ExoSound	2.69	2.06	2.56
ExoVisual	1.31	1.69	1.31

Table 19: Mean ratings

Ease of Use

Table 20 shows that in terms of Ease of Use, there was a statistically significant difference in the response to this ranking between the conditions, $\chi^2(3) = 31.350$, $p < 0.001$. A post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .0083$. There was a statistically significant difference in the ranking between the Egocentric + Visual cue and Egocentric + Sound cue conditions (Z=-3.404, p=.001), the Exocentric + Sound cue and Egocentric + Sound cue conditions (Z=-2.917, p=.004), the Exocentric + Visual cue and Egocentric + Sound cue conditions (Z=-3.617, $p < .001$), and the Exocentric + Visual cue and Exocentric + Sound cue conditions (Z=-3.380, p=.001).

Subjects found that the Exocentric + Visual cue method was the easiest to use, while the Egocentric + Sound cue conditions was the least easy to use.

Ease of Use Ranking: Pairwise tests						
	EgoVis - EgoSnd	ExoSnd - EgoSnd	ExoVis - EgoSnd	ExoSnd - EgoVis	ExoVis - EgoVis	ExoVis - ExoSnd
Z	-3.404	-2.917	-3.617	-1.407	-2.018	-3.380
Asymp. Sig.	.001	.004	.000	.159	.044	.001

Table 20: Wilcoxon signed ranks test results

Fun

Table 21 shows that in terms of Fun, there was a statistically significant difference in the response to this ranking between the conditions, $\chi^2(3) = 18.375$, p <0.001. A post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at p <.0083. There was a statistically significant difference in the ranking between the Exocentric + Sound cue and Egocentric + Sound cue conditions (Z=-3.005, p=.003) and the Exocentric + Visual cue and Egocentric + Sound cue conditions (Z=-2.675, p=.007).

Subjects found that when the Sound cue method was used, the Exocentric interface made the tasks more fun. The Exocentric + Visual cue condition was also perceived as more fun than the Egocentric + Sound cue condition.

Fun Ranking: Pairwise tests						
	EgoVis - EgoSnd	ExoSnd - EgoSnd	ExoVis - EgoSnd	ExoSnd - EgoVis	ExoVis - EgoVis	ExoVis - ExoSnd
Z	-1.931	-3.005	-2.675	-1.615	-2.372	-1.324
Asymp. Sig.	.053	.003	.007	.106	.018	.185

Table 21: Wilcoxon signed ranks test results

Effectiveness

Table 22 shows that in terms of Effectiveness, there was a statistically significant difference in the response to this ranking between the conditions, $\chi^2(3)$ = 25.725, p <0.001. A post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at p <.0083. There was a statistically significant difference in the ranking between the Exocentric + Sound cue and Egocentric + Sound cue conditions (Z=-2.665, p=.0076), the Exocentric + Visual cue and Egocentric + Sound cue

conditions (Z=-3.516, p<.001) and the Exocentric + Visual cue + Exocentric + Sound cue conditions (Z=-3.346, p=.001) .

Subjects found the Exocentric + Visual cue condition the most effective.

Effectiveness Ranking: Pairwise tests						
	EgoVis - EgoSnd	ExoSnd - EgoSnd	ExoVis - EgoSnd	ExoSnd - EgoVis	ExoVis - EgoVis	ExoVis - ExoSnd
Z	-2.507	-2.665	-3.516	-.054	-2.338	-3.346
Asymp. Sig.	.012	.008	.000	.957	.019	.001

Table 22: Wilcoxon signed ranks test results

Summary

Generally the Exocentric + Visual cue method scores the best on many questions. Comparing the two cuing methods, the Visual cue method was perceived better than the Sound cue method. When comparing the interfaces, the Exocentric tablet interface received better scores on the questionnaire than the Egocentric tablet interface.

5.3.4 Qualitative Measures

At the end of each questionnaire subjects were asked to write down some comments they might have. They were asked to briefly describe how the system made them feel, and what was or was not intuitive. People generally answered that they liked the Exocentric + Visual cue condition the best and that it was more easy, efficient and intuitive than other conditions.

Please briefly explain how the system made you feel.

In the Egocentric + Sound cue condition, about 30% of the subjects reported that it was hard and annoying to localize the sounds. However, after using it for a while, subjects learned the voices of the people and got more used to the system. At that point they could enjoy the tasks. One subject said *"I could rarely identify what direction the sound came from. I used the voices. When I recognized the voice the system was easy and fun"*. Another issue was that subjects found it confusing and hard to switch their focus from the HMD view to the tablet view, and back to the HMD view. This caused disorientation in some subjects. One subject reported that in this case, the requested file had to be memorized, while this wasn't necessary for the Visual cue conditions.

The Egocentric + Visual cue condition partly received the same comments as the Sound cue variant. Here, subjects also found it confusing to change their focus from the HMD to the tablet and back again. Another interesting comment was that because of the absence of sound or moving avatars the virtual people on the HMD seemed very fake or dead, making the subject feel uncomfortable and disconnected from the scene. This particular person said: *"It made me feel disconnected because there was no interaction from the participants in the Virtual Environment, only the artificial shapes on top of their faces. They seemed dead."*

The Exocentric + Sound cue condition received similar comments as that of the Egocentric + Sound cue condition, with subjects saying that it was relatively hard to localize the sounds. However subjects say that it was rewarding when they did find the correct person on the HMD. They also found it easier to use than the Egocentric conditions, because the radar view was easier to interpret and gave more information on their surroundings.

Subjects' comments show that the Exocentric + Visual cue condition feels the most fun, efficient and satisfying. One subject noted that it felt like a game. However, some people commented that the HMD is redundant in this condition, since there is no real need to look on the HMD as visual cues also appear on the tablet. It also felt like a traditional drag and drop action, making subjects feel confident and comfortable when doing the action. Some people said that they thought they were cheating by just using the tablet and said things like *"I felt like I was doing it wrong but it was also nice and easy."*.

Please briefly explain what was intuitive and/or not intuitive in the system

For the Egocentric + Sound cue condition, subjects were mostly bothered by looking repeatedly up and down from the HMD view to the tablet view, and found this unintuitive. However, looking around in the 3D space did feel intuitive for most subjects. One subject noted that the selection method of the virtual people by looking at them felt intuitive in the beginning, but selection was difficult to maintain when trying to send a file. They said *"Selecting people by looking at them seems intuitive at the beginning, but it was hard to look at two screens at the same time."*

While there were similar concerns about having to switch between the use of the two displays for the Egocentric + Visual cue condition, subjects felt this

condition was more intuitive and logical. Some subjects commented that a visual cue would also be less disruptive in a real conference call. Subjects also felt that the identification of the visual cues was easier than that of the sound cues. The 'share with all' cue was slightly confusing, since it appeared on a different location than the other cues.

Subjects found the tablet interface in the Exocentric + Sound cue condition very intuitive, and liked how the screen updated depending on the direction in which they were looking. Dragging the file on top of the avatar also made sense and felt intuitive. In this case, dragging to the centre to share with everybody also felt natural for subjects.

The Exocentric + Visual cue method mostly received positive comments and was regarded as very intuitive. However, having the HMD but not having to use it sometimes felt unintuitive and weird. People recognized the kind of action they had to do and felt comfortable using a touch screen in this way, making it very intuitive.

5.3.5 Observations

Observations made by the researcher during the experiments are very similar to comments made by the subjects themselves in the previous section. Subjects often seemed confused then using the Sound cues. Especially in the Egocentric + Sound cue condition they had to look around a lot using the HMD, and often did not get what target was speaking, even if they were close to looking at the right one.

During the Exocentric + Visual cue condition subjects often seemed confused for a little bit, as the HMD was not necessary to complete the tasks. Many asked if it was allowed to just use the HHD, and said they felt they were cheating if they did. As example, one subject said *"Ehm.. is it OK if I just use the tablet?"*. Many others said similar things.

For some subjects, the headphones and HMD were giving them obvious discomfort on the ears and nose. The discomfort grew as the experiment continued. For one subject the gear had to be taken off in between every condition.

In general, subjects enjoyed the experiment, as it used technology that was new to most of them, such as the monocular HMD. However, most of them also found wearing the HMD and headphones quite heavy and irritating after some

time. Also some of their movement was hindered by the cables that connected the HMD and headphones to the PC.

5.4 Conclusions

Overall the results show that performance is generally faster and better when using the Exocentric + Visual cue condition. Performance mostly increased when the sound cues were replaced by visual cues. When the subject was using the visual cues, the performance time and head movement was significantly better when the Exocentric interface was used instead of the Egocentric interface. The questionnaire results also showed significantly decreased confusion or disorientation, increased perceived simplicity and performance, increased enjoyment and increased satisfaction.

Subjects generally performed the worst when using the Egocentric + Sound cue condition. It also causes slight confusion and disorientation. The sound cues were not always recognized as easily as visual cues. However, subjects commented that they found it easier when they started to recognize what voice belonged to what person.

Both hypotheses stated at the start of this chapter have to be rejected. Most measurements showed a significant difference between the two interfaces, with the Exocentric interface performing significantly better than the Egocentric. This means that H1 has to be rejected. H2 has to be rejected because significant differences have been found between the use of the sound and visual cues. Overall the visual cues performed better. In the next chapter we discuss these results in more detail, and in particular describe some design guidelines that can be learned from them that might be helpful for developers building similar interfaces.

6 Discussion and Design Guidelines

This chapter provides interpretations of the results found in the user study. Afterwards it will also list a set of proposed design guidelines for AR conferencing applications using HHD-HMD interfaces.

6.1 Discussion

In the user study we found that users were able to complete the task fastest in the Exocentric + Visual cue condition. This is likely because of the fact that participants had no need for the HMD when using this interface with the visual cues. Since these cues also appeared on the HHD, and the participant did not have to look around to send a file, all actions could be completed by looking down at the HHD, ignoring the HMD. This resulted in significantly decreased head movement and time. This is also supported by the questionnaire results that showed significantly decreased confusion and disorientation, increased perceived simplicity and performance, increased enjoyment and increased satisfaction.

The Egocentric + Sound cue condition performed worst, probably due to two reasons. First, this interface requires a user to physically look around, which takes time and might also be prone to accuracy errors. It also causes slight confusion and disorientation. Second, the sound cues were not always recognized as easily as visual cues. Finding the right direction that the sound came from proved to be more difficult for participants than expected. Participants had to listen carefully and to remember the requested file, while the visual cue was persistent and did not have to be remembered. However, participants also commented that they found it easier to use the sound cues when they started to recognize which voice belonged to what person.

Another possible explanation for the increased performance with visual cues is the fact that there is no interpretation needed of the shapes and colours. In the sound cue conditions, the words have to be recognized and translated into the concept of a shape and colour. In the visual cue conditions however, this is unnecessary as the file is already represented by a visual shape and colour.

It is likely that the Exocentric interface performed better because it did not require the user to look around as much as the Egocentric interface. It was also less prone to accuracy errors as selection was done on the HHD with the fingers, instead of the HMD with head movements. Users also did not need to change their focus from the HMD to HHD and back as often as in the Exocentric interface. This resulted in a different user experience flow. For the Egocentric

interface, the flow was as follows, with a total of four steps, of which two require focus on both the HMD and the HHD simultaneously:

1. Find target on HMD (focus: HMD)
2. Find file on HHD (focus: HHD)
3. Hold finger on HHD and look back up to the target on HMD (focus: HHD-HMD)
4. Swipe up on HHD without while keeping the target selected on HMD (focus: HHD-HMD)

However, when using the Exocentric interface, the flow is reduced from four steps to three, with no need for shared focus:

1. Find target on HMD (focus: HMD)
2. Find file and target on HHD (focus: HHD)
3. Swipe file to target on HHD (focus: HHD)

There were elements in the prototype used for the user study that could be improved if the study was to be repeated. For instance, in the egocentric interface, user were expecting a swipe gesture to move the data files, while in reality the application was reacting to a drag and drop gesture. When the user swiped up to send a file, a swipe gesture could still be recognized by the application *if* the user's finger slid off the screen at the top of the screen. This is recognized as a drop because the finger is no longer touching the screen. However, when swiping down, this did not work as the Android control bar is between the lower edge of the application and the lower edge of the touch area. Thus, a user had to do a proper drag and drop into the lower area, which was relatively small. These shortcomings have to be taken in consideration as they might have had influence on the final results. Improving these elements might give different results.

6.2 Design Guidelines

Based on the results of the user study the following and the discussion made in the previous section, design guidelines were developed that could be used by people creating AR conferencing applications using HHD-HMD systems. These capture the lessons learned about the influence of different cuing methods and interfaces for a HHD-HMD system. The proposed guidelines are listed below:

1. *Use spatial audio to give an initial sense of direction, confirm a user's expectation with a visual cue.*

   While spatial audio can add realism and a feeling of immersion to an AR conferencing environment, it also has a different effect on different people. According to the results from this research, spatial audio cannot be used as a reliable cuing method for file exchange in an AR conferencing space. In a case were people ask for files, a user will preliminary recognize the voice rather than its spatial origin. To show which conference participants need files, visual cues seem to be understood more easily by users. Spatial sound cues can be used to give an initial sense of direction, but users will need a visual confirmation in order to feel confident enough to take action.

2. *Use an exocentric interface when a user is not moving and has to interact with elements placed around the user. Consider an egocentric interface when a user is moving and in a game-like environment.*

   In a mobile AR conferencing application, an egocentric tablet interface might work in certain scenarios, but an exocentric interface proved much easier to understand for users. The Exocentric interface gives a clear overview of the conference situation and shows what the user expects to see. An egocentric tablet interface requires a user to move around a lot and with that lose accuracy and time for doing the actual task. In a situation where the user is not moving around, an egocentric interface may cause confusion and decreases task efficiency. However, in a game-like environment where challenge is important and a user also moves to different locations, it might be possible to apply it.

3. *Only switch the focus between HMD and HHD when necessary. Give certainty to the user that it is safe and useful to switch focus.*

   In most applications users will not be able to focus on both the HMD and HHD at the same time. They will also find it uncomfortable to switch their focus between the displays too often and fast. Both displays exist in a different space for the user and switching between those spaces for no good reason can be disorientating. For instance, in our egocentric condition, users had to keep looking at the target to select it, but also had to glance down at the tablet interface to chose the right file and drag it up.

Some participants had to repeatedly look up and down to check if they were doing the correct action. It seems that not just the head movement is causing the extra confusion, but also the mental switching between the spaces. Either a switch from the main display to an alternative display should be short, returning the focus to the main display within moments, or the focus kept on the alternative display until an action requires the focus to switch back to the main (head-mounted) display.

4. *Require the user's attention only on one task at a time.*

   Tying in with the previous guideline, users struggle concentrating on two tasks simultaneously. In the case of our study, users had to do two tasks that involved moving two body parts independently from each other at the same time (move their head to focus on the HMD target, and their finger to send the file). For most people, this is a generally a hard thing to accomplish and requires some training. Completing two tasks at the same time that do not involve movement might be easier, but to avoid a steep learning curve of an application it is good practice to require a user's attention only on one task at a time.

5. *Use an HHD for precise gestures, and head tracked HMD for looking around and rough gestures.*

   The results of our study show that the Exocentric HHD interface performed significantly better. One factor that is likely to have caused this is the precise interaction users had to perform to send a file. In case of the Exocentric interface, users could use their finger to drag and drop a file, which is a task that requires some accuracy but is easily achieved. In the Egocentric interface however, users had to precisely move their head to target the right virtual conference participant. This required more time and head movements, generally decreasing performance. However, users did enjoy being able to look around with the head tracked HMD.

7 Conclusion and Future Work

7.1 Conclusion

Augmented Reality superimposes information on the real world, which gives it the potential to be used as a remote conferencing tool, transforming every location into a conferencing space. AR and head mounted displays can be used to support workers in remote locations with live annotations that overlay the real world, or by giving heads-up warning signals and other information. AR has also been integrated in numerous mobile phone applications, such as games or navigation aids. AR on head mounted displays has been combined with hand held devices to open up alternative interaction methods. However, there are few clear guidelines or resources to rely on when designing AR applications for HMD and HHD, especially for remote collaboration.

As discussed in the introduction, our research aimed to contribute to the field of mobile AR conferencing by exploring different interfaces and cuing methods, and to create a basic set of design guidelines that can be used for future work. This has been achieved by designing a prototype application that emulates an AR conferencing space with file-sharing capabilities. After a needs analysis and brainstorming process, two HHD interfaces and cuing methods were designed. Using these designs prototype systems were created using off-the-shelf components.

The two interfaces compared were egocentric and exocentric, both offering the user a different way of sending files to conference participants. In the egocentric case users use the HMD to look at a conference participant before flicking a file their way. In the exocentric case users could use a top-down radar-like view of the virtual conference space to directly drop files on conference participants.

The two cuing methods compared were audio and visual. For the audio cues, users heard spatial audio requests that helped them orientate in the 3D virtual space. For the visual cues, users saw representations of the requested file on the HMD and HHD.

After developing the prototypes a formal user study was conducted to see which method was best for file transfer between AR conference participants. In our evaluation sixteen participants tested the four conditions. We found that an exocentric interface is easier for users to understand and generally improves performance over an egocentric interface. Using spatial audio proved less effective

than visual cues. When using the exocentric interface, visual cues significantly improved performance. Both H1 and H2 had to be rejected, as significant differences were found between the two interfaces and also between the two cuing methods.

The study of the two HHD interfaces contributed to the research area by laying a foundation for future work about AR file sharing applications using HMD-HHD systems. The research also showed that there was a significant interaction between the HHD interface and the cuing method used. Future research and applications could take these results in mind when designing interfaces. The results of this study and its discussion led to the creation of a set of design guidelines for future research and application development for HMD and HHD mobile AR conferencing applications. These include:

- Use spatial audio to give an initial sense of direction, confirm a user's expectation with a visual cue.
- Use an exocentric interface when a user is not moving and has to interact with elements placed around the user. Consider an egocentric interface when a user is moving and in a game-like environment.
- Only switch the focus between HMD and HHD when necessary. Give certainty to the user that it is safe and useful to switch focus.
- Require the user's attention only on one task at a time.
- Use an HHD for precise gestures, and head tracked HMD for looking around and rough gestures.

7.2 Future Work

Our research has just begun to explore what can be done with HMD and HHD for AR conferencing. There are still many things left to explore that are outside the scope of this thesis. It showed the current opportunities of the combined use of HMD and HHD for AR conferencing. It inspires us to improve the current interfaces and apply similar interaction techniques to other applications. The technology used in this research is constantly improving, which opens up more and more possibilities for future research in this area. In this section we describe interesting areas for future work and future technology that will be useful in this area of research and design.

7.2.1 Interfaces

In this research we compared two tablet interfaces and tested audio and visual cuing methods and how they interacted with each other. However, we have not tested a system were both cuing methods were combined. A combination of sound and visual cues could lead to better results for both tablet interfaces. For example, spatial audio can be used to give an initial sense of direction, and visual cues can be used to confirm the users expectation. Audio cues might also work better with other sounds than voices, such as beep notifications. In this case the perceived spatiality of the sound could also be improved for instance by changing the pitch or frequency of beeps. This creates opportunities of future research in which more cuing methods and combinations could be tested. It would also be interesting to see what other spatial cues can be used for an AR conferencing space. We discussed spatial audio and visual cues, but other options include haptic cues or other audio types (pitch, amplitude). For example the haptic feedback element in the HHD could vibrate more intensly when looking in the direction of the target. An HHD with one haptic element both on the left and right could be used to give a sense of direction by vibrating one of them more intensely than the other.

The prototype system could still be improved in many aspects. Based on the user feedback the usability of the egocentric interface could be improved. For instance, the lower 'share publicly' drop area is, in retrospect, badly designed. A better way to share publicly would be to use the HMD to look at an empty space or the ground, and then use the upper tablet area to swipe it to the HMD. This eliminates the sudden change of interaction between public and private shares.

7.2.2 Applications

An important area of future work is to test similar systems with real conference participants instead of simulated ones. This would create a more natural environment for users and might give different results. This could be set up by having at least three people remotely work together, on a joint task such as shared construction. This task could be created so that one user is 'on location' and is asked to do the constructing, while the others are remote users who have parts of a construction manual. To create an incentive for file sharing, there could be multiple construction tasks and manuals. A construction task could then be randomly picked at the beginning of the experiment, encouraging participants to share photo's of the construction task before starting. This would be a more realistic testing environment compared to our study for testing the us-

ability of the whole system. It is likely that people will feel more involved when communicating with real people than when using our simulated conference.

It will be also interesting to test the same interfaces (egocentric and exocentric) in a different application, for instance gaming. In our research we have seen that egocentric interfaces are not suitable when a user is not expected to move around but forced to look around. However, in a more mobile application where are user is already moving around the use of the egocentric interface could show different results. Or maybe a single user application rather than a shared conferencing application could be tested.

7.2.3 Head Mounted Displays

There is also future work that could be done with different hardware, especially new types of Head Mounted Displays. HMDs are currently entering the consumer market and high quality devices are becoming very affordable. For example, products such as the Oculus Rift [32], shown in Figure 45 have recently started shipping for game developers and enthusiasts for only $300 USD.

Figure 45: The Oculus Rift Head Mounted Display

The Oculus Rift has a wide 110 degree field of view of 110 degrees and a fast 9 degrees of freedom head tracker. Developers can also get access to the SDK and examples of the integration in game engines such as the Unreal Engine and Unity. So far users have been overwhelmed by the immersive Virtual Reality experience of the Rift. It is not an optical see-through device, however one could easily modify the device to use stereo cameras, turning it into a video see-through AR display. This device is perfect for creating extremely immersive

Virtual (or when modified, Augmented) reality applications.

A completely different kind of HMD, but also noteworthy is Google Glass (see Figure 46) [16]. Google Glass was first revealed to the public in early 2012 and has started to ship prototypes to developers and other individuals since early 2013. The display features a compact design with a small glass display just above the eye, not unlike the Brother Airscouter. The display resolution is 640 by 360 pixels. It has a small bone conductive speaker that is situated just behind the ear that sends vibrations directly through the bone to produce sound. For input methods it uses speech recognition and a small touch enabled surface on the side of the device for touch and swipe commands. It can also be connected to external input devices such as Bluetooth keyboards or touch pads. Current prototypes are connected to the internet using WiFi or via a Bluetooth connection with a phone. The display is in the piece of glass above the eye, which functions mostly as a Heads-up Display, providing the wearer with information whenever needed. It would definitely be an interesting piece of hardware to tailor our application to.

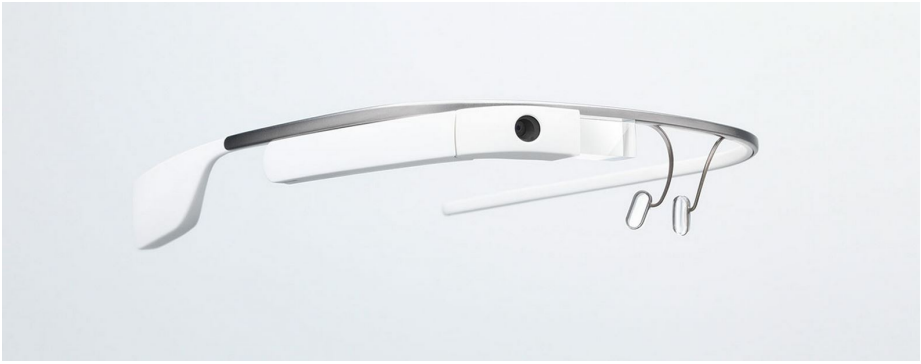


Figure 46: Google Glass

7.2.4 Concept Design for Google Glass

We would like to conclude this thesis with a future concept design for a virtual conference using Google Glass and an HHD interface. This concept is a redesigned version of the application created for our research to make it more suitable for Google Glass.

In the application voice commands are used to start a conference call with a selected group of people, for instance by saying *"Conference with Cecile, James and Timo"*. The call is then placed and the video feeds of the selected users

appear (see Figure 47). Instead of using a 3D virtual space, the interface shows a 2D strip of video feeds. A user can rotate their head just slightly to the left or right to change what person they are looking at. A possible option is to automatically show the current speaker in the centre of the display. As soon as the call is initiated, voice commands are no longer preferred since the other conference participants might get annoyed or confused by the user speaking to the device. From this point on, the HHD could be used for interaction.

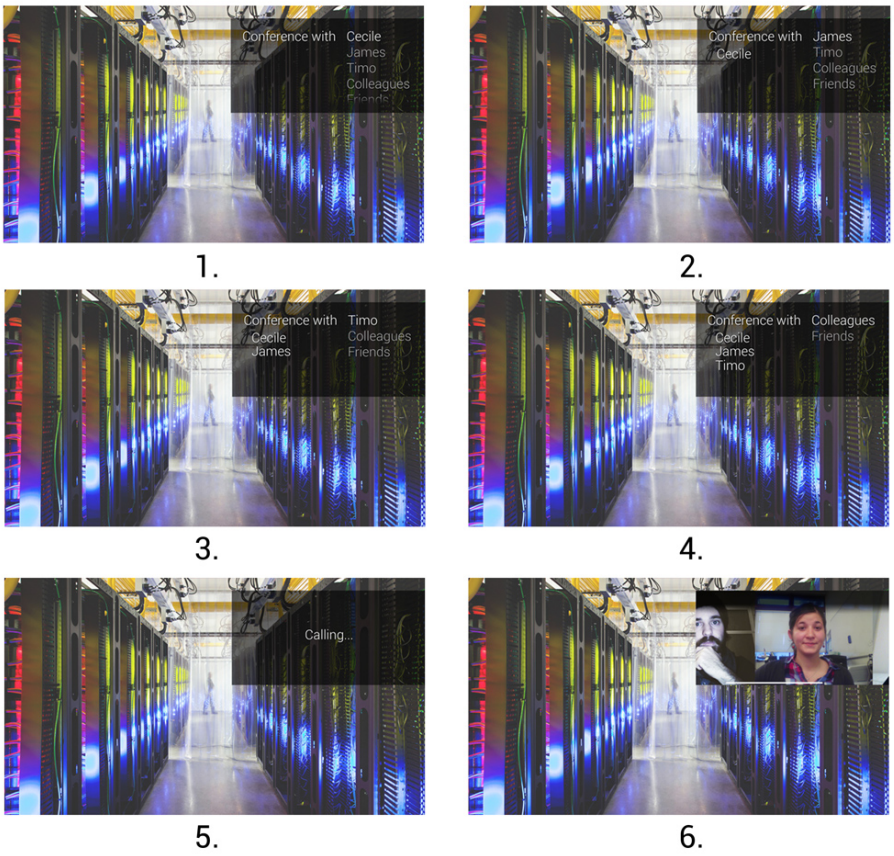


Figure 47: Interacion design concept for Google Glass

When a user requests a file, they can of course do so by asking the user. Additionally, they can send a cue to the user. As Google Glass only has a single speaker, spatial audio is harder to implement than with a stereo headset. However, other sound cues could be used, as discussed before. A visual cue would slide down from the top of the screen. The text shows who requested the file and is displayed in a higher opacity when focusing on the requester (see Figure 48). On the HHD, the user will see a simple representation of the conference.

This is visualized as half a circle with the participants evenly spread out over the shape. When a user focuses on a participant with Google Glass, the area under their avatar on the HHD changes colour to indicate selection. There will always be one person selected, so its not necessary to precisely focus. Moreover, the HHD will use a sending method similar to that of our tested exocentric interface, where a user could drag the item onto a participant's avatar. This also helps eliminating the need for precise focusing on conference participants. A scrollable list of items is shown below the half circle. This can show the files that have been received and sent. Further design would be needed to support a better file browsing experience. However, in this concept the focus is on file sharing and not on file browsing.

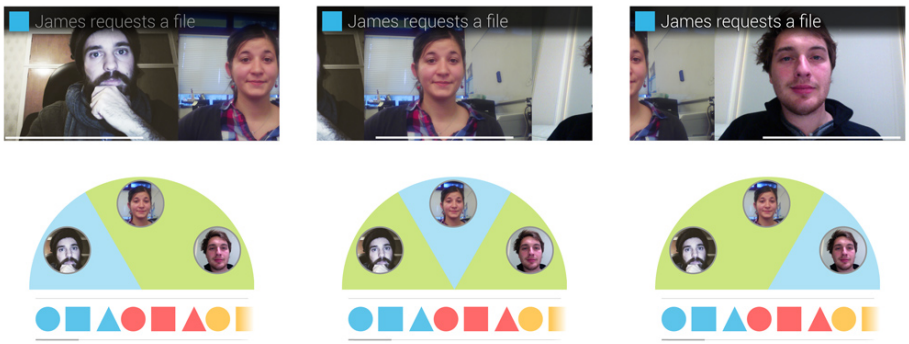


Figure 48: Interaction design concept for Google Glass (upper 3 images) and HHD (lower 3 images)

In summary, new devices such as Google Glass will provide more immersive and more wearable AR experiences. These in turn will enable new types of remote collaboration applications. Our work has shown that HMDs and HHDs can be used together for effectively for remote conferencing, but there are more opportunities that need to be explored in order to make the technology mainstream.

8 References

References

[1] Jessica J. Baldis. Effects of spatial audio on memory, comprehension, and preference during desktop conferences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '01, pages 166–173, New York, NY, USA, 2001. ACM.

[2] M. Billinghurst, J. Bowskill, M. Jessop, and J. Morphett. A wearable spatial conferencing space. In *Wearable Computers, 1998. Digest of Papers. Second International Symposium on*, pages 76–83. IEEE, 1998.

[3] A. Butz, T. Hollerer, C. Beshers, S. Feiner, and B. McIntyre. An experimental hybrid user interface for collaboration. Technical report, DTIC Document, 2005.

[4] H.H. Clark and S.E. Brennan. Grounding in communication. *Perspectives on socially shared cognition*, 13(1991):127–149, 1991.

[5] Christina Dicke, Shaleen Deo, Mark Billinghurst, Nathan Adams, and Juha Lehikoinen. Experiments in mobile spatial audio-conferencing: key-based and gesture-based interaction. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*, MobileHCI '08, pages 91–100, New York, NY, USA, 2008. ACM.

[6] Andreas Dünser, Raphael Grasset, and Mark Billinghurst. A survey of evaluation techniques used in augmented reality studies. In *ACM SIGGRAPH ASIA 2008 courses*, pages 5:1–5:27, New York, NY, USA, 2008. ACM.

[7] M.R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):32–64, 1995.

[8] S. Feiner, B. MacIntyre, T. Höllerer, and A. Webster. A touring machine: Prototyping 3d mobile augmented reality systems for exploring the urban environment. *Personal and Ubiquitous Computing*, 1(4):208–217, 1997.

[9] S.R. Fussell, L.D. Setlock, J. Yang, J. Ou, E. Mauer, and A.D.I. Kramer. Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction*, 19(3):273–309, 2004.

[10] Taejin Ha and Woontack Woo. Arwand: Phone-based 3d object manipulation in augmented reality environment. In *Ubiquitous Virtual Reality (ISUVR), 2011 International Symposium on*, pages 44–47. IEEE, 2011.

[11] J. Hauber, H. Regenbrecht, A. Hills, A. Cockburn, and M. Billinghurst. Social presence in two-and three-dimensional videoconferencing. 2005.

[12] Jörg Hauber, Holger Regenbrecht, Mark Billinghurst, and Andy Cockburn. Spatiality in videoconferencing: trade-offs between efficiency and social presence. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 413–422. ACM, 2006.

[13] Steven J Henderson and Steven Feiner. Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*, pages 135–144. IEEE, 2009.

[14] A. Hills, J. Hauber, and H. Regenbrecht. Videos in space: a study on presence in video mediating communication systems. In *ACM International Conference Proceeding Series*, volume 157, pages 247–248, 2005.

[15] Tobias Höllerer, Steven Feiner, Tachio Terauchi, Gus Rashid, and Drexel Hallaway. Exploring mars: developing indoor and outdoor user interfaces to a mobile augmented reality system. *Computers & Graphics*, 23(6):779–785, 1999.

[16] Google Inc. Google glass. `http://www.google.com/glass/start`, June 2013.

[17] S. Izadi, A. Agarwal, A. Criminisi, J. Winn, A. Blake, and A. Fitzgibbon. C-slate: a multi-touch and object recognition system for remote collaboration using horizontal surfaces. In *Horizontal Interactive Human-Computer Systems, 2007. TABLETOP'07. Second Annual IEEE International Workshop on*, pages 3–10. IEEE, 2007.

[18] R.E. Kraut, S.R. Fussell, and J. Siegel. Visual information as a conversational resource in collaborative physical tasks. *Human-computer interaction*, 18(1):13–49, 2003.

[19] R.E. Kraut, M.D. Miller, and J. Siegel. Collaboration in performance of physical tasks: Effects on outcomes and communication. In *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, pages 57–66. ACM, 1996.

[20] James R Lewis. Ibm computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7(1):57–78, 1995.

[21] The Center For New Music and Audio Technology. Opensoundcontrol. `http://opensoundcontrol.org/introduction-osc`, June 2013.

[22] openFrameworks.cc. openframeworks. `http://www.openframeworks.cc`, June 2013.

[23] D. Palmer, M. Adcock, J. Smith, M. Hutchins, C. Gunn, D. Stevenson, and K. Taylor. Annotating with light for remote guidance. In *Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces*, pages 103–110. ACM, 2007.

[24] Ronald Poelman, Oytun Akman, Stephan G. Lukosch, and Pieter Jonker. As if being there: Mediated reality for crime scene investigation. In *CSCW '12:Proceedings of the ACM 2012 conference on Computer supported cooperative work*. ACM, 2012.

[25] Chandrasekhar Ramakrishnan. Javaosc. `https://www.illposed.com/software/javaosc.html`, February 2013.

[26] Gerhard Reitmayr and Dieter Schmalstieg. Collaborative augmented reality for outdoor navigation and information browsing. In *Proc. Symposium Location Based Services and TeleCartography*, pages 31–41, 2004.

[27] D. Schmalstieg, A. Fuhrmann, and G. Hesina. Bridging multiple user interface dimensions with augmented reality. In *Augmented Reality, 2000.(ISAR 2000). Proceedings. IEEE and ACM International Symposium on*, pages 20–29. IEEE, 2000.

[28] Ivan E Sutherland. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pages 757–764. ACM, 1968.

[29] Zsolt Szalavári and Michael Gervautz. The personal interaction panel–a two-handed interface for augmented reality. In *Computer Graphics Forum*, volume 16, pages C335–C346. Wiley Online Library, 1997.

[30] Firelight Technologies. Fmod. `http://fmod.org/index.html`, June 2013.

[31] Martin Usoh, Ernest Catena, Sima Arman, and Mel Slater. Using presence questionnaires in reality. *Presence: Teleoperators & Virtual Environments*, 9(5):497–503, 2000.

[32] Oculus VR. Oculus rift. `http://www.oculusvr.com`, June 2013.

[33] Bob G Witmer and Michael J Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998.

[34] Michelle Yeh, James L Merlo, Christopher D Wickens, and David L Brandenburg. Head up versus head down: The costs of imprecision, unreliability, and visual clutter on cue effectiveness for display signaling. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45(3):390–407, 2003.

9 Appendix

9.1 Appendix A: Consent Form

The following information and consent form were given to participants before the start of the study.

INFORMATION FORM

RESEARCH STUDY: Interaction Design for the combined use of a Heads-up Display and Hand Held Display and Spatial Audio for a 3D augmented conferencing space.

INVESTIGATORS: Timo Bleeker, Prof Mark Billinghurst, Dr. Gun Lee, Dr Andreas Dünser.

You are invited to participate in the research study entitled: Interaction Design for the combined use of a Heads-up Display and Hand Held Display and Spatial Audio for a 3D augmented conferencing space.

The aim of this project is to investigate different user interfaces and interaction methods for sharing files in a 3D augmented reality environment

Your participation in this experiment will have you perform a series of tasks across multiple interfaces. The tasks will include listening for audio cues and sharing files with virtual people. Between each interface you will be given a short questionnaire to rate your usage of each interface. After all interfaces have been used a final questionnaire will be given to rate your overall impression.

You may, at any time request to withdraw from the experiment for any reason with no consequence and your participation in the experiment will be terminated.

Upon the completion of your involvement in this study, we will also provide you with a $5 gift voucher.

The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation: the identity of the participants will not be made public without their consent. To ensure anonymity and confidentiality, only the researchers will be allowed access to the video recordings of the participants. The recordings will be destroyed after a period of 5 years. They will be put in a secure and encrypted location that requires a password to gain access. The only individuals with the password will be the researchers in the study. Furthermore, the collected research data will also be kept in a secure and locked location. Only the researchers will have access to it via a key.

The project is being carried out as a research project by Timo Bleeker, under the supervision of Prof. Mark Billinghurst and Dr. Gun Lee, who can be contacted by the following means. They will be pleased to discuss any concerns you may have about participation in the project.

Timo Bleeker
HIT Lab NZ, University of Canterbury
Email: timo.bleeker@pg.canterbury.ac.nz

Prof Mark Billinghurst
HIT Lab NZ, University of Canterbury
Email: mark.billinghurst@canterbury.ac.nz

Dr. Gun Lee
HIT Lab NZ, University of Canterbury
Email: gun.lee@hitlabnz.org

Dr Andreas Dünser.
HIT Lab NZ, University of Canterbury
Email: andreas.duenser@hitlabnz.org

This proposal has been reviewed and approved by the Human Interface Technology Laboratory, University of Canterbury and the University of Canterbury Human Ethics Committee Low Risk process.

Please take this form with you when you leave.

CONSENT FORM

RESEARCH STUDY: Interaction Design for the combined use of a Heads-up Display and Hand Held Display and Spatial Audio for a 3D augmented conferencing space.

INVESTIGATORS:

Timo Bleeker
HIT Lab NZ, University of Canterbury
Email: timo.bleeker@pg.canterbury.ac.nz

Prof. Mark Billinghurst
HIT Lab NZ, University of Canterbury
Email: mark.billinghurst@canterbury.ac.nz

Dr. Gun Lee
HIT Lab NZ, University of Canterbury
Email: gun.lee@hitlabnz.org

Dr Andreas Dünser.
HIT Lab NZ, University of Canterbury
Email: andreas.duenser@hitlabnz.org

I have read and understood the description of the above-named project. On this basis I agree to participate voluntarily as a subject in the project, and I consent to publication of the results of the project with the understanding that anonymity will be preserved. The data will be kept for up to 5 years before being destroyed.

I understand that my actions will be recorded during the experiment, but the recording will only be viewed by researchers directly associated with the project. I also understand that the recorded data will be kept for up to 5 years before being destroyed.

I understand also that I may at any time withdraw from the project, including withdrawal of any information I have provided.

I note that the project has been reviewed ***and approved*** by the University of Canterbury Human Ethics Committee.

_____	_____	_____
Participant (Print name)	Signature	Date

9.2 Appendix B: Questionnaire

Participants were asked to fill in the following questionnaire. Some of the pages were repeated and asked after each condition.

Participant #:

ARoom: Augmented Conference and File-sharing Room.

Thank you for participating in our research. Please take some time to fill in the first part of this questionnaire before we begin the trial. If you have any questions, please ask us.

About You

1. How old are you?

   _____ years old
2. What is your Gender?
   - Male
   - Female
3. Do you have experience with augmented reality and virtual reality?
   - Yes, I use either AR or VR regularly.
   - Yes, I have used AR and VR some times in the past
   - No, I never used any AR or VR applications
4. Do you have experience using tablets?
   - Yes, I use a tablet daily
   - Yes, I sometimes use a tablet
   - No, I never use tablets
5. Are you colour-blind? If yes, what type of colour-blindness do you have?
   - Yes, type: _____
   - No

Participant #:

The following questions give you an opportunity to tell us your experience with the system you just used. Your responses will help us understand what aspects of the system you are particularly concerned about and the aspects that satisfy you.

Please read each statement and indicate how strongly you agree or disagree with the statement by ticking the corresponding box on the scale.

Participant #:

Condition Name

Condition #:	Disagree	<	<>	>	Agree
Overall, I am satisfied with how easy it is to use this system.					
It was simple to use this system.					
I could effectively complete the tasks and scenarios using this system.					
I felt comfortable using this system.					
It was easy to learn to use this system.					
I believe I could become productive quickly using this system.					
The interface of this system was pleasant.					
I liked using the interface of this system.					
Overall, I am satisfied with this system.					
The interactions with the environment seemed natural.					
I was able to anticipate what would happen in response to the actions that I performed.					
I could easily identify the cues. (The icons or sound that appeared)					
I could easily localize the cues. (The icons or sound that appeared)					
I felt confused or disorientated at the end of the session.					
I was felt involved in the virtual environment.					
I experienced delay between my actions and the expected outcomes.					
In the end I became proficient in interacting with the virtual environment.					

Participant #:

Please briefly explain how the system made you feel and why. (Annoyed, happy, stressed, etc.)

Please briefly explain what was intuitive and/or not intuitive in the system.

Participant #:

Please rate all the conditions from high (1) to low (4).

	Sound + Radar	Sound + Swipe	Visual + Radar	Visual + Swipe
Fun				
Easy to use				
Most Effective				

Thank you for your cooperation with this research. If you have any other comments please write them down here. If you would like to contact us later, please refer to the Information Form you have received at the start of the research.